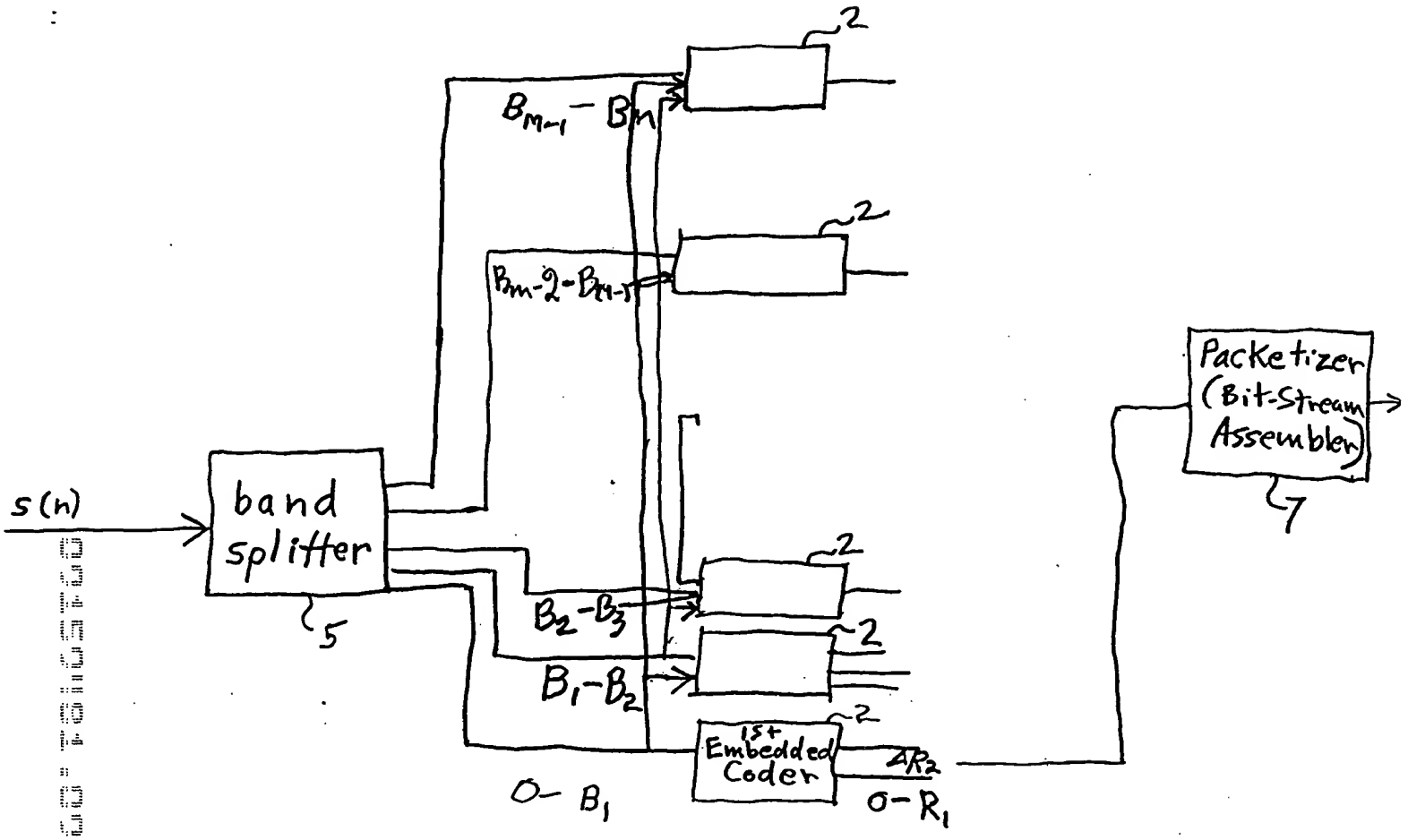

2
$B_{M-1} - B_M$
2
$B_{M-2} - B_{M-3}$
Packetizer (Bit-Stream Assembler)
7
s(n)
band splitter
5
2
$B_2 - B_3$
2
$B_1 - B_2$
2
1st Embedded Coder
ΔR_2
$O - B_1$
$O - R_1$

Fig. 1A

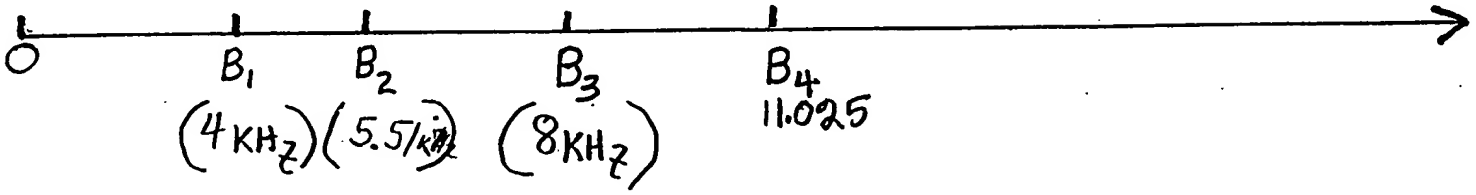

O
B_1
(4 KHz)
B_2
(5.5 KHz)
B_3
(8 KHz)
B_4
11.025

Fig. 1B

FFT-based Scalable and Embedded Codec Architecture

—— Encoder with M Octave Bands

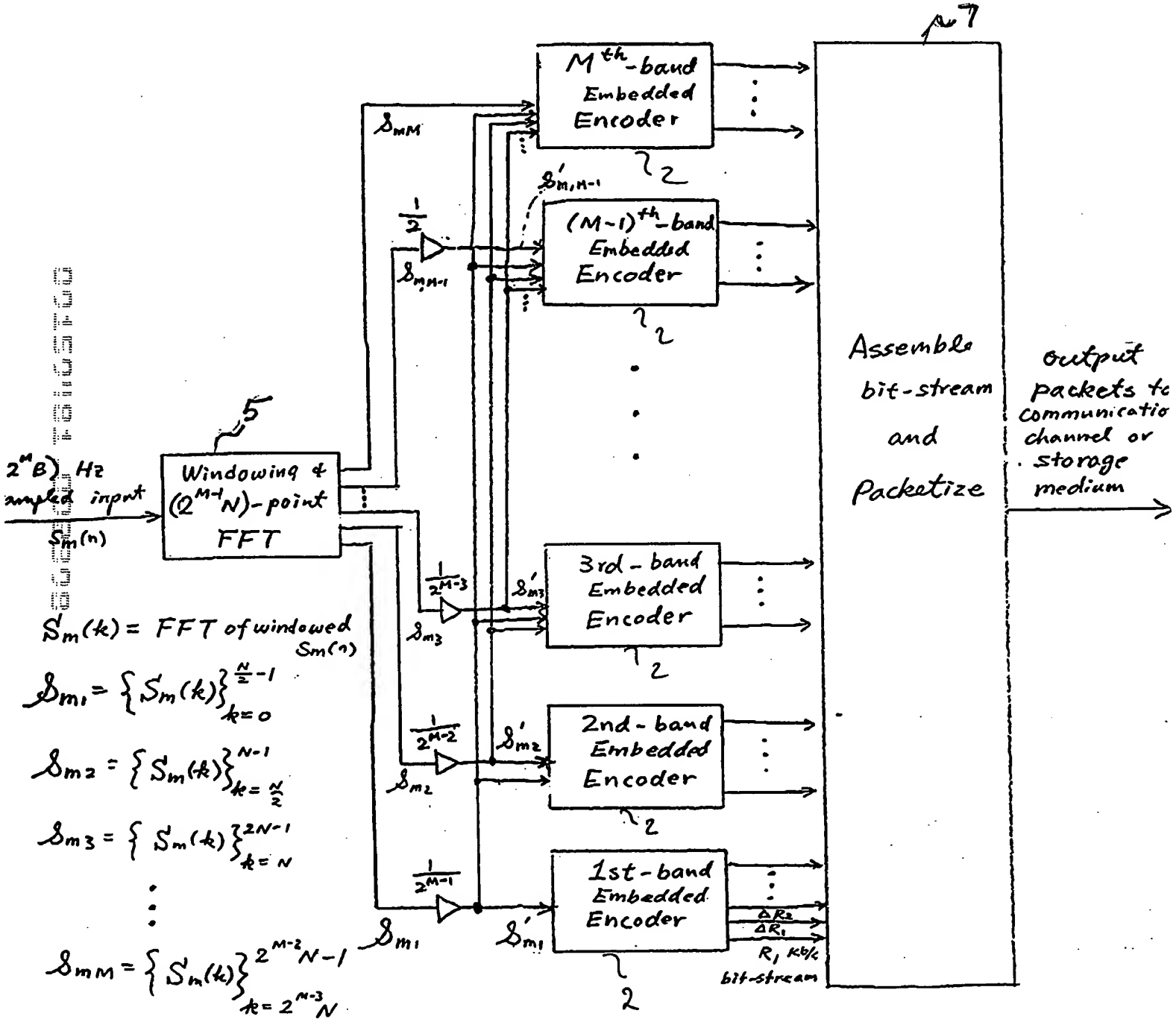


Fig. 2A

FFT-based Scalable and Embedded Codec Architecture

— Decoder with M_1 Octave Bands ($1 \le M_1 \le M$)

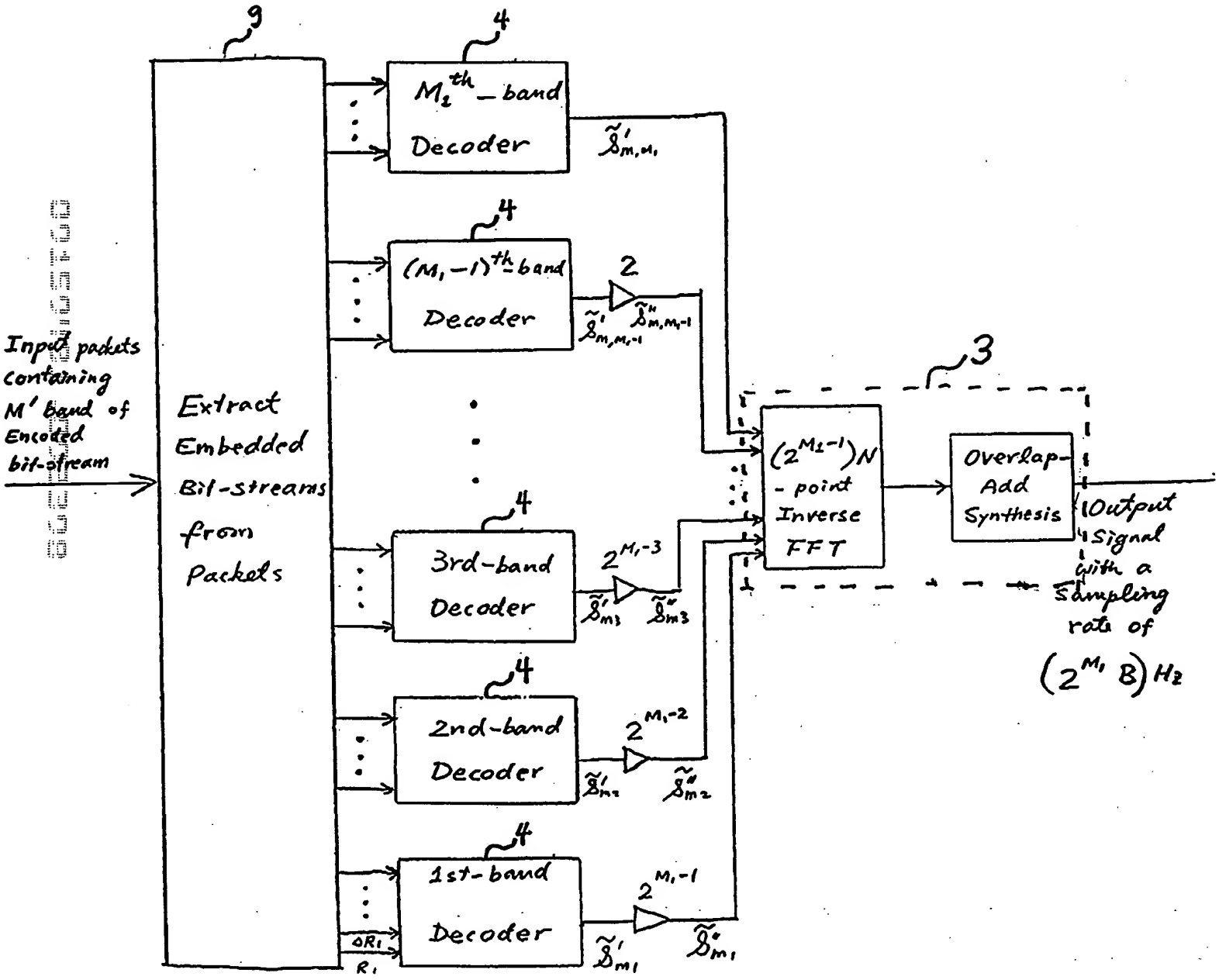


Fig. 2B

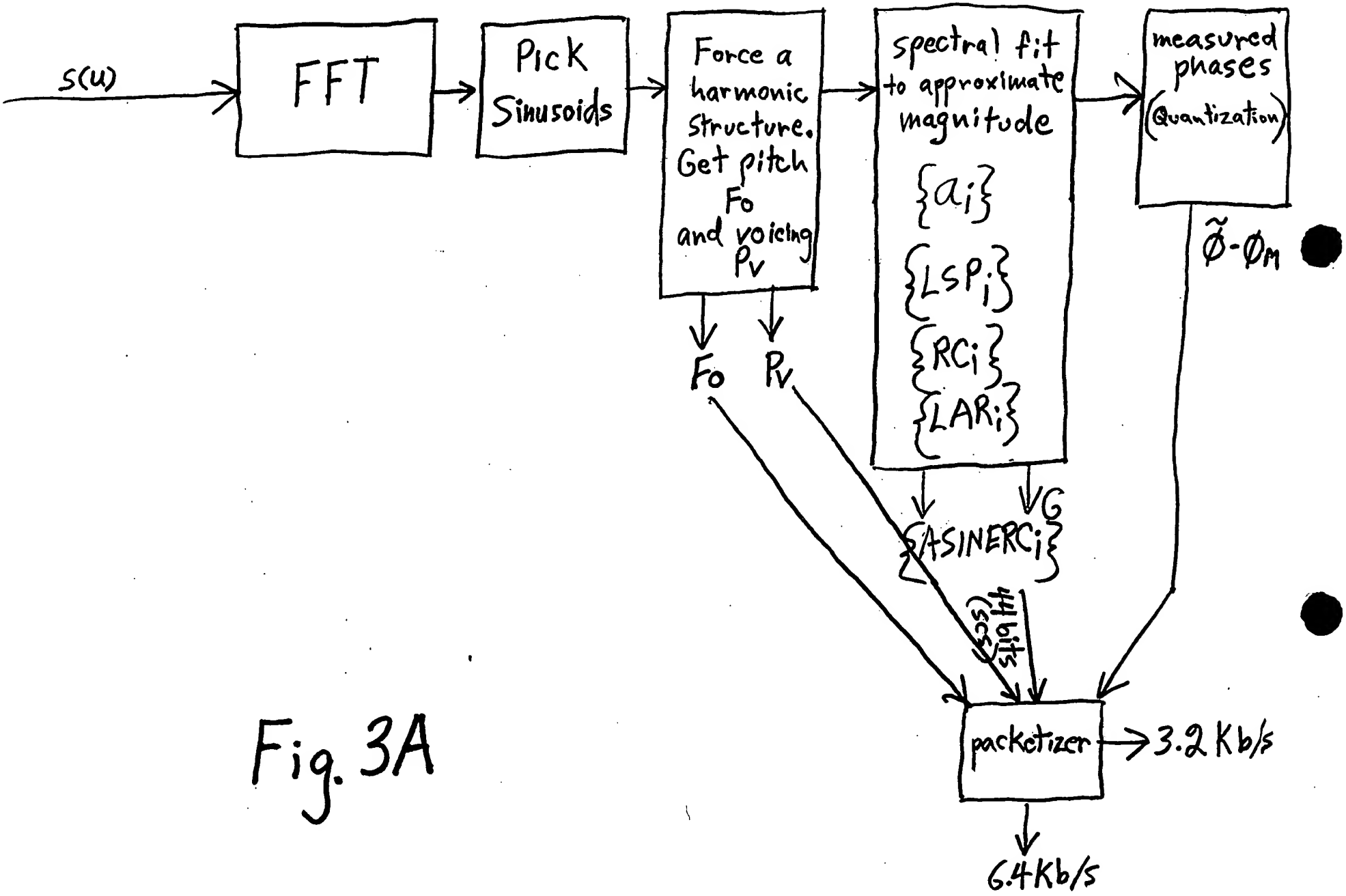
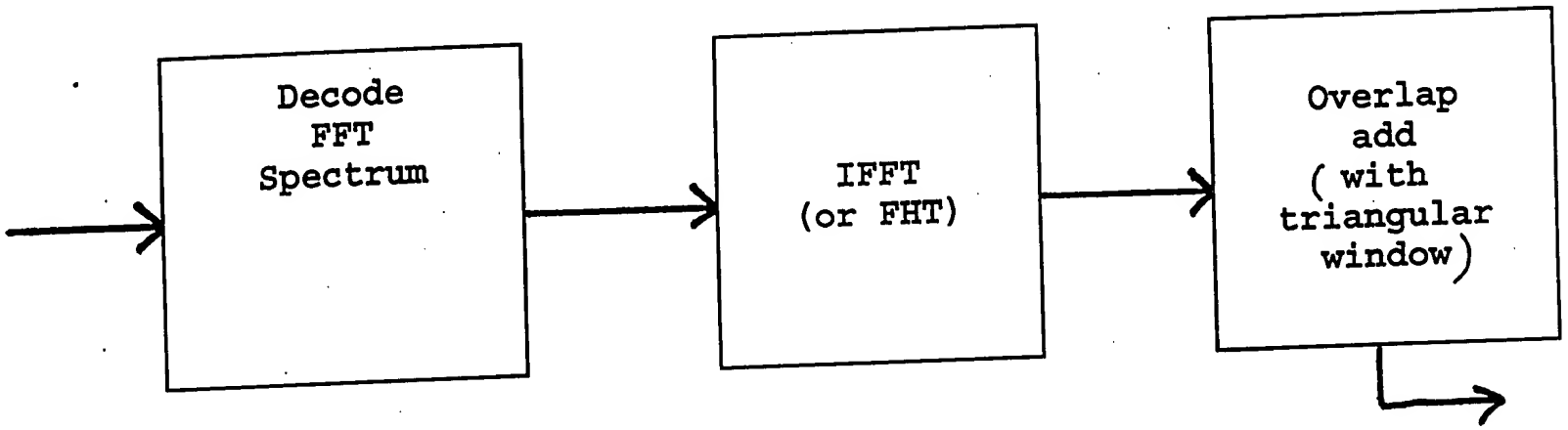

s(u)
FFT
Pick Sinusoids
Force a harmonic structure. Get pitch Fo and voicing Pv
spectral fit to approximate magnitude
{ai}
{LSPi}
{RCi}
{LARi}
measured phases (Quantization)
Φ̃-ΦM
Fo
Pv
G
{ΔSINERCi}
44 bits (SCS)
packetizer
3.2 Kb/s
6.4Kb/s

Fig. 3A



Decoder: Synthesis every M ms.

Fig. 3B

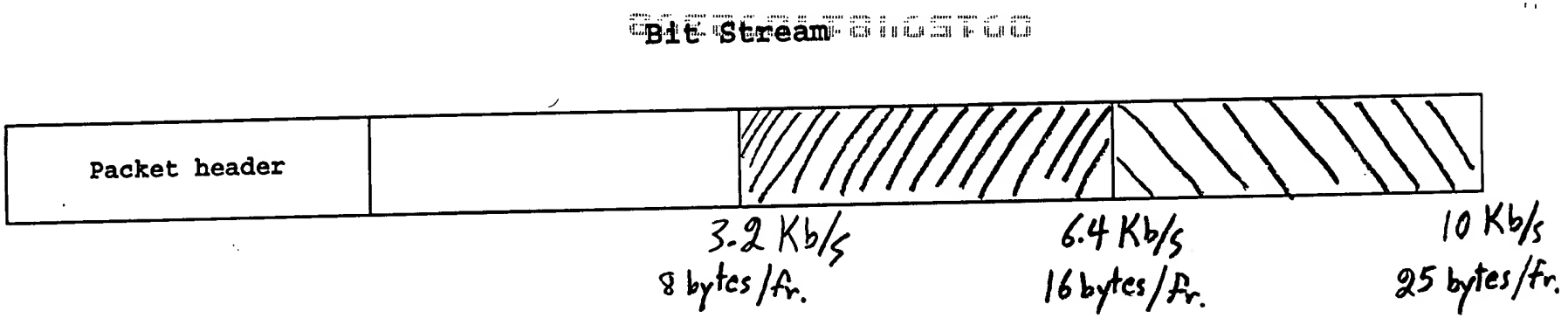

Bit Stream
Packet header
3.2 Kb/s
8 bytes/fr.
6.4 Kb/s
16 bytes/fr.
10 Kb/s
25 bytes/fr.

Fig. 4A

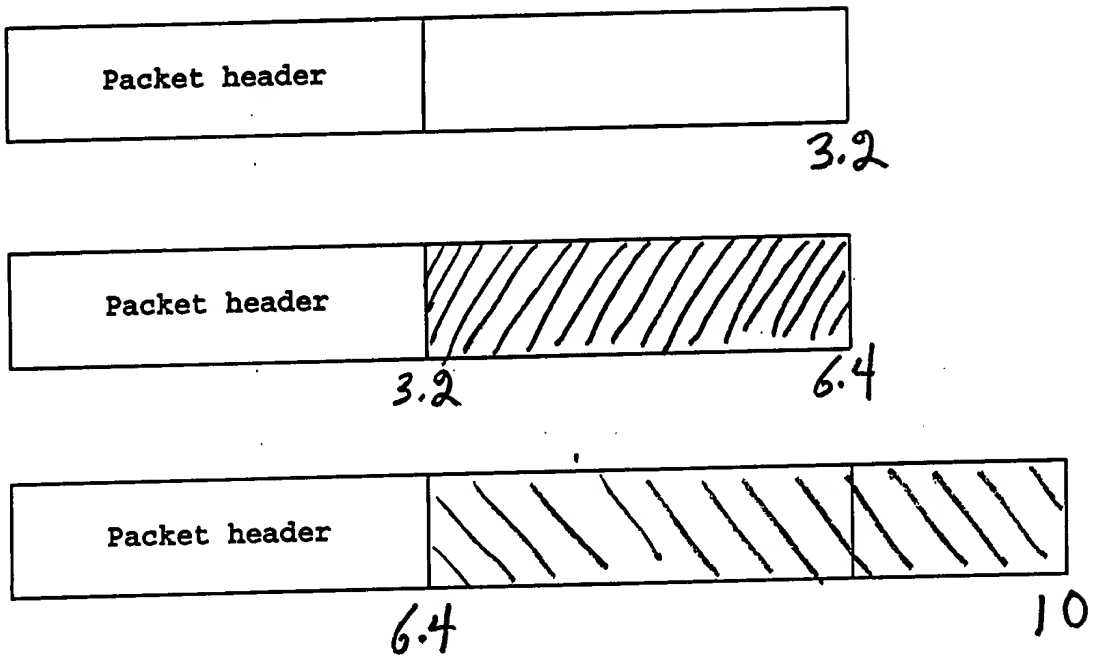

Packet header
3.2
1st priority packet
Packet header
3.2
6.4
2nd priority packet
Packet header
6.4
10
3rd priority packet

Fig. 4B

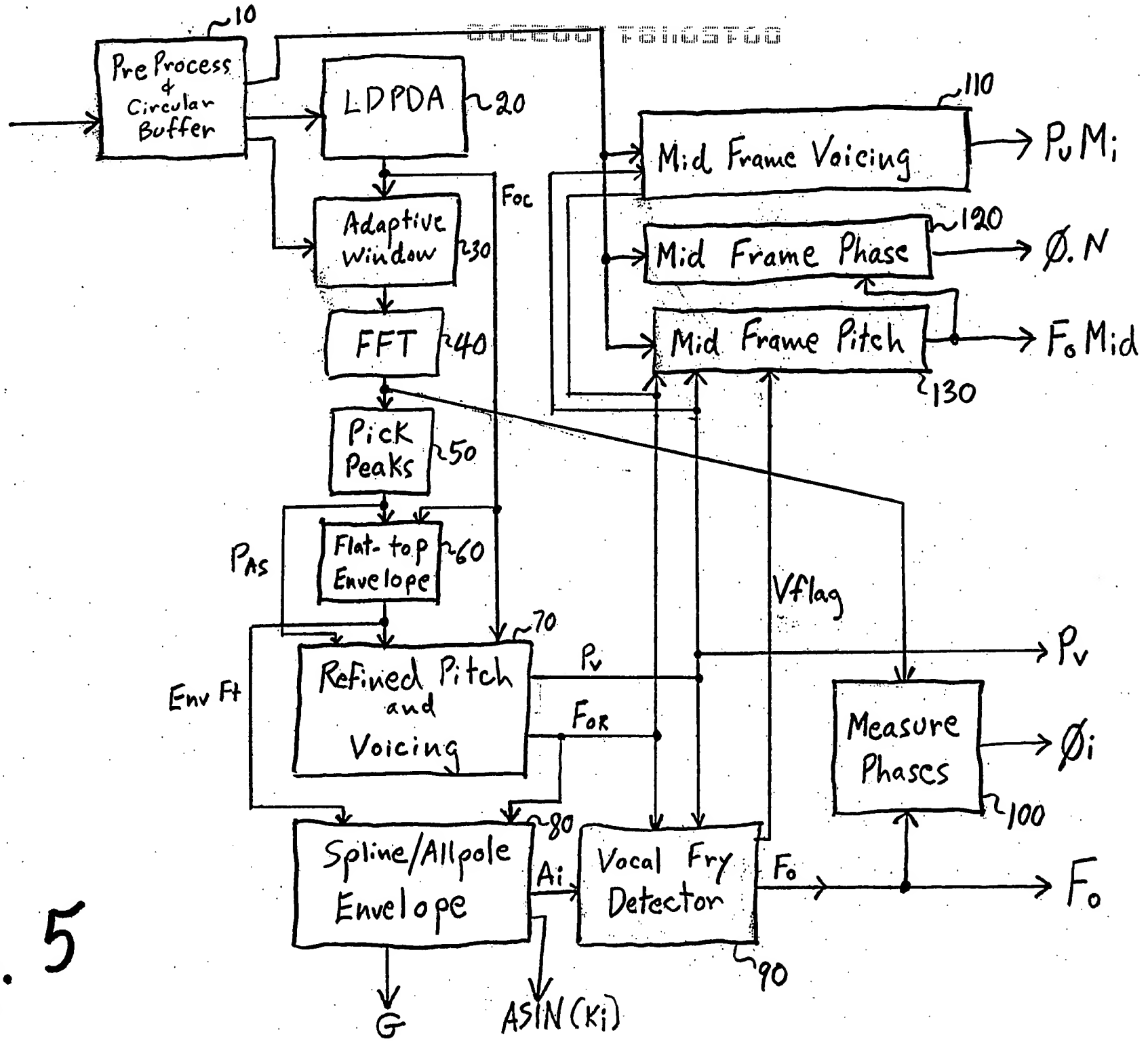

10
Pre Process & Circular Buffer
LDPDA
20
Foc
Adaptive Window
30
FFT
40
Pick Peaks
50
PAS
Flat-top Envelope
60
70
Env Ft
Refined Pitch and Voicing
Pv
FoR
80
Spline/Allpole Envelope
Ai
G
ASIN(Ki)
Vocal Fry Detector
90
Fo
Vflag
110
Mid Frame Voicing
PvMi
120
Mid Frame Phase
Ø.N
Mid Frame Pitch
130
FoMid
Measure Phases
100
Øi
Pv
Fo

Fig. 5

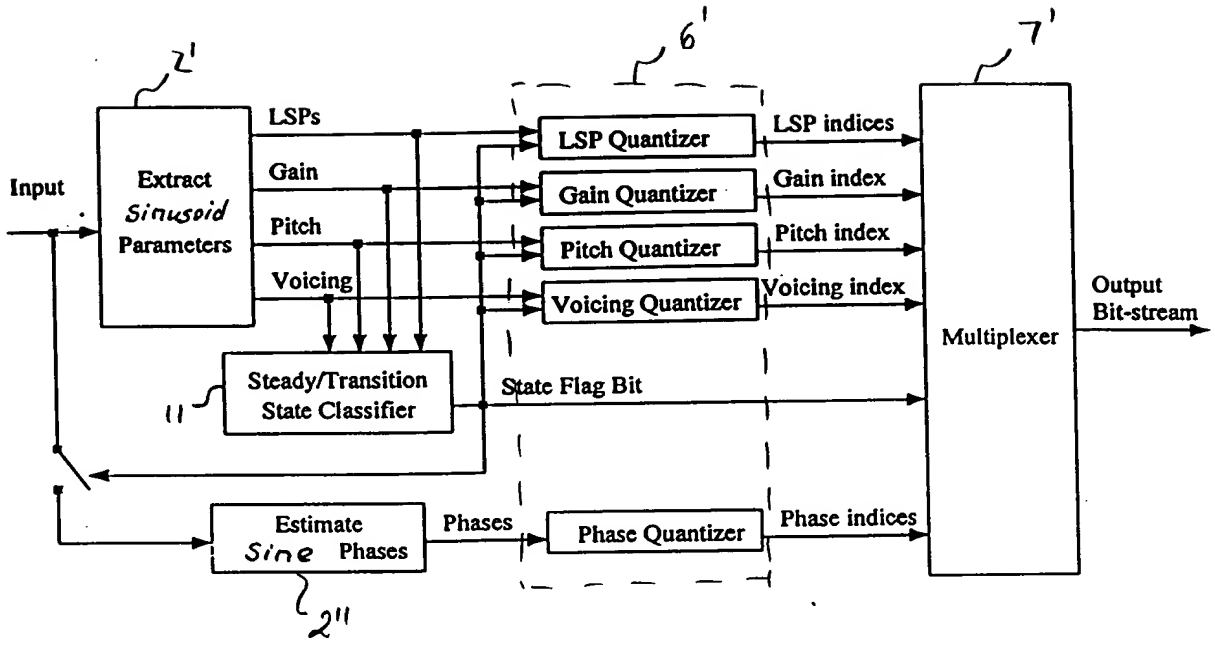

2'
6'
7'
Input
Extract
Sinusoid
Parameters
LSPs
Gain
Pitch
Voicing
LSP Quantizer
Gain Quantizer
Pitch Quantizer
Voicing Quantizer
LSP indices
Gain index
Pitch index
Voicing index
Multiplexer
Output
Bit-stream
Steady/Transition
State Classifier
11
State Flag Bit
Estimate
Sine Phases
Phases
Phase Quantizer
Phase indices
21

Fig 5A

SPEECH SYNTHESIZER

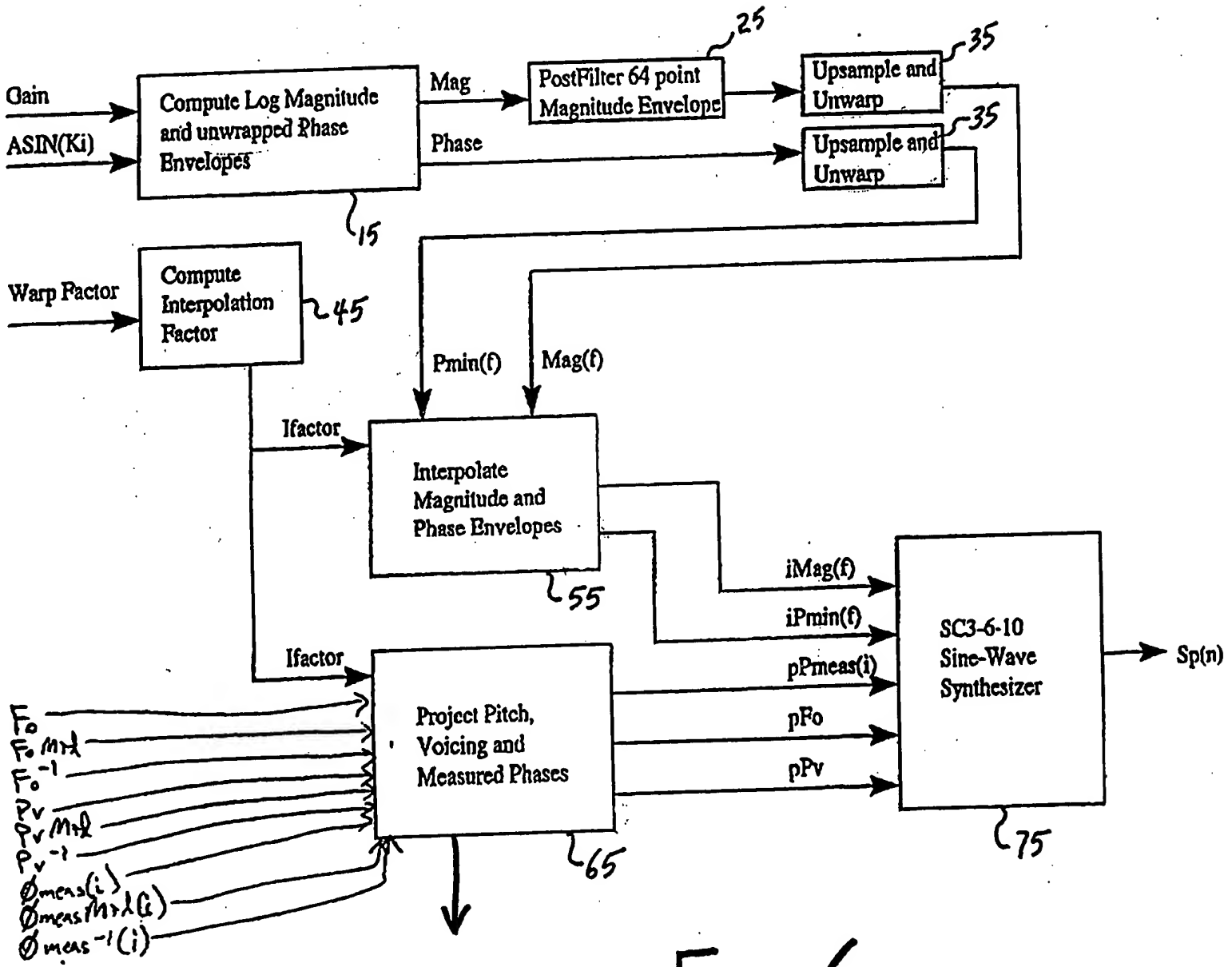


Fig. 6

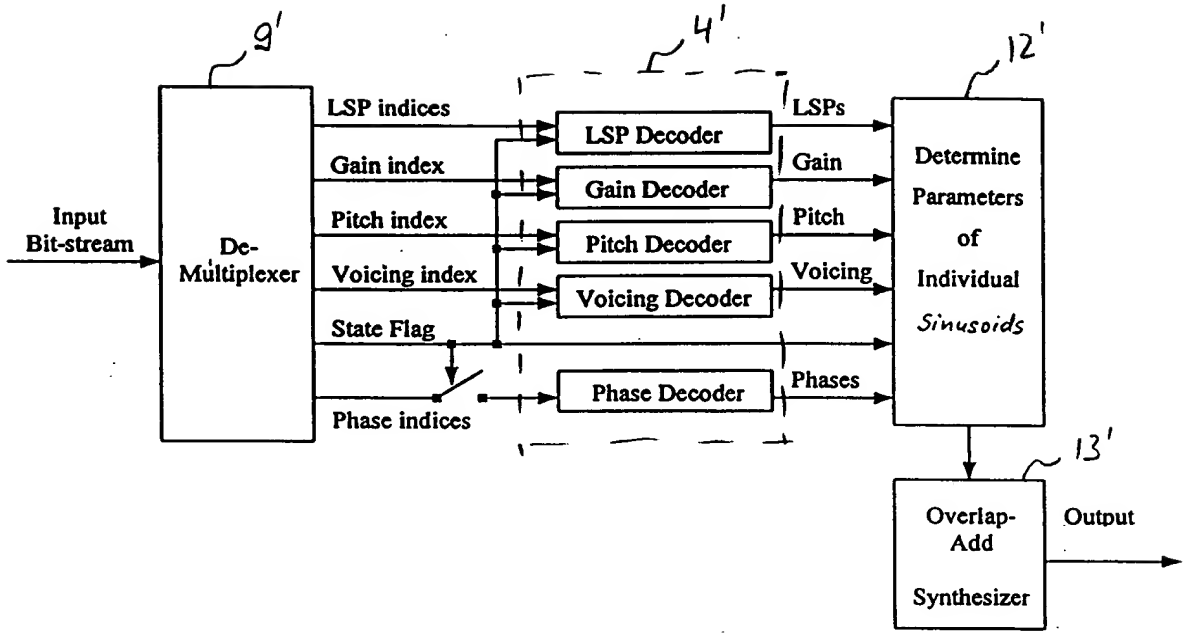
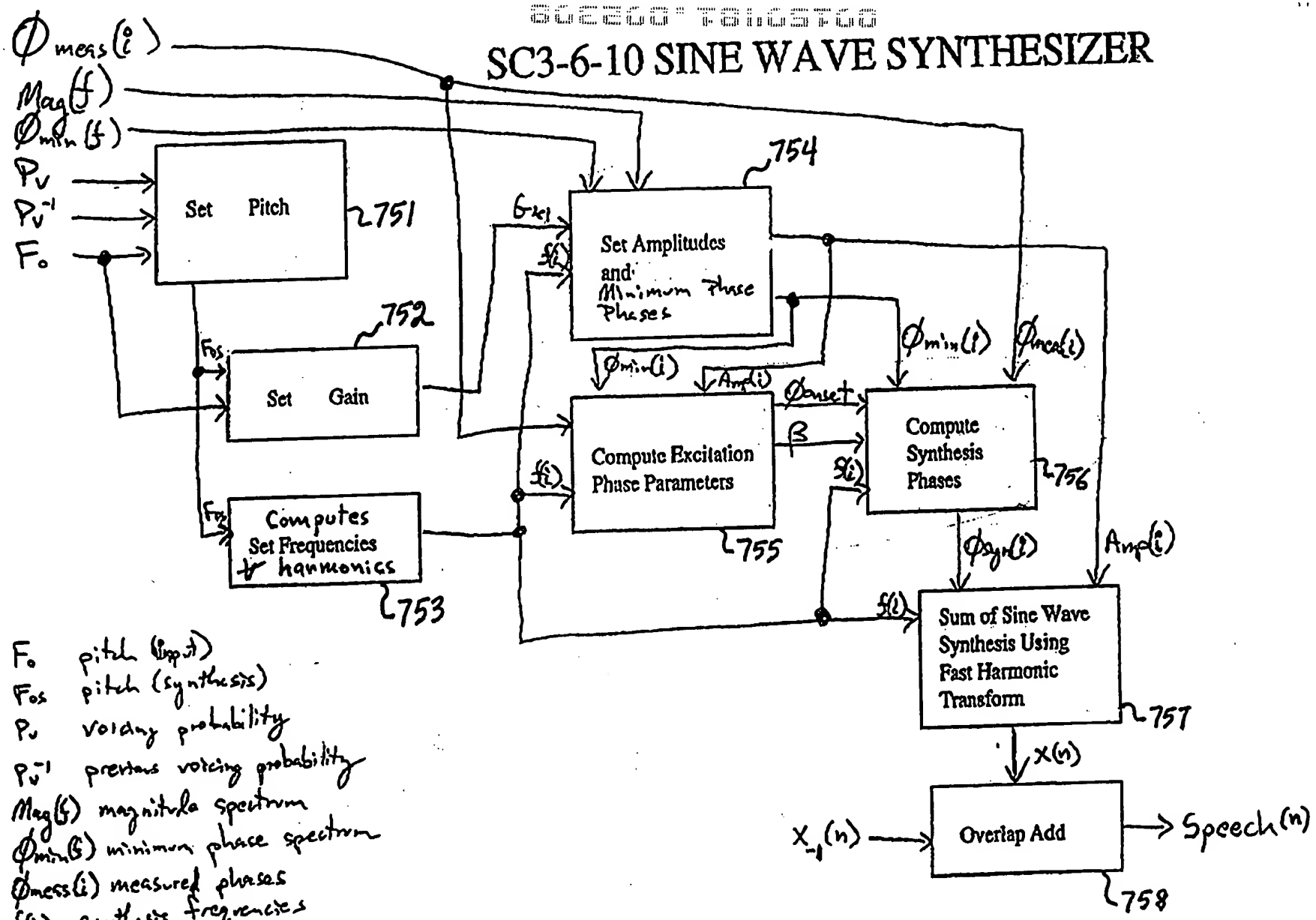

9'
4'
12'
13'
Input Bit-stream
De-Multiplexer
LSP indices
Gain index
Pitch index
Voicing index
State Flag
Phase indices
LSP Decoder
Gain Decoder
Pitch Decoder
Voicing Decoder
Phase Decoder
LSPs
Gain
Pitch
Voicing
Phases
Determine Parameters of Individual Sinusoids
Overlap-Add Synthesizer
Output

Fig. 6A

SC3-6-10 SINE WAVE SYNTHESIZER



F_o pitch (input)
F_{os} pitch (synthesis)
P_v voicing probability
P_v^{-1} previous voicing probability
Mag(f) magnitude spectrum
$\phi_{min}(f)$ minimum phase spectrum
$\phi_{meas}(i)$ measured phases
f(i) synthesis frequencies
Amp(i) synthesis amplitudes
$\phi_{syn}(i)$ synthesis phases
$\phi_{min}(i)$ minimum phases at f(i)

Fig. 7

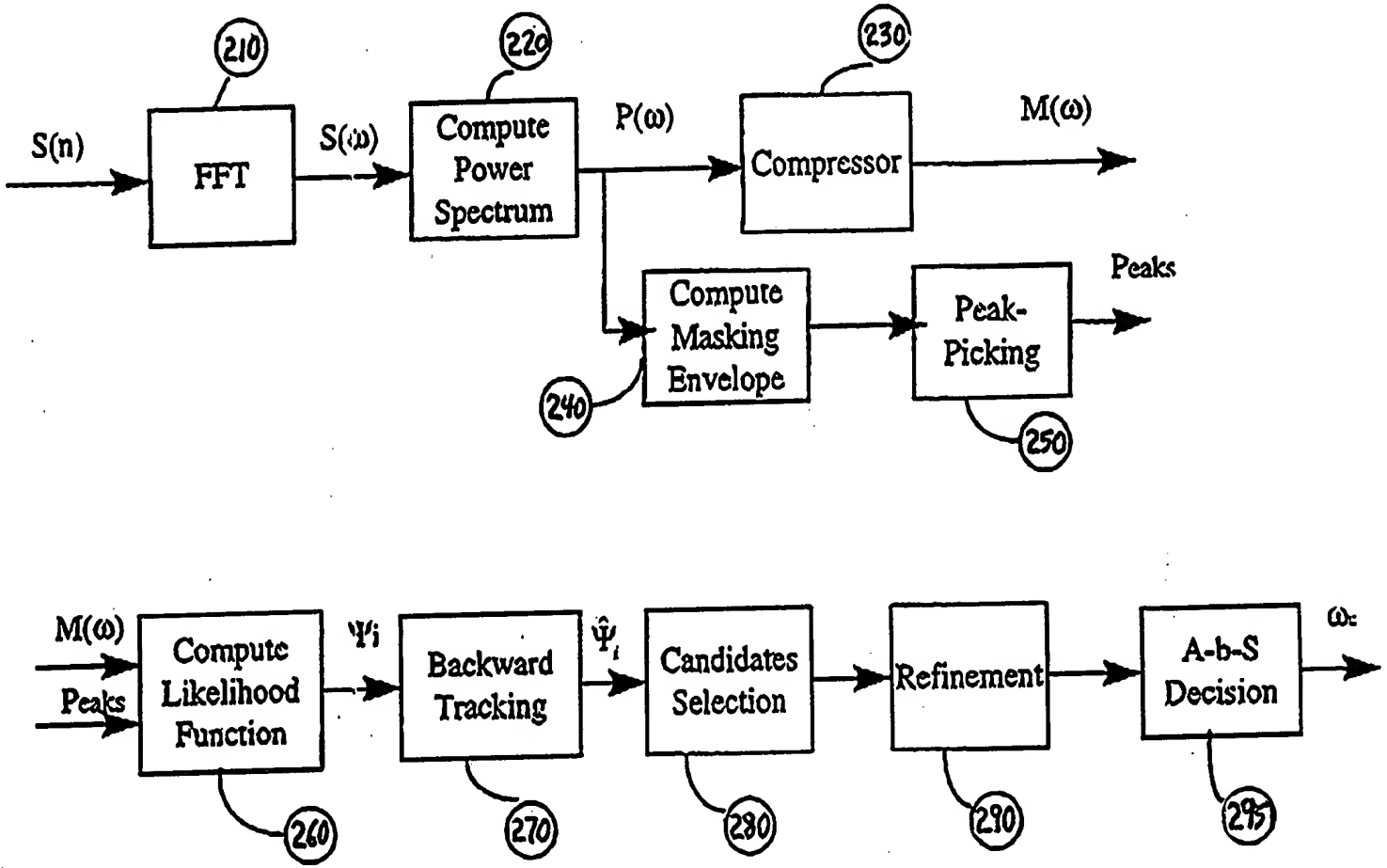


Fig. 8

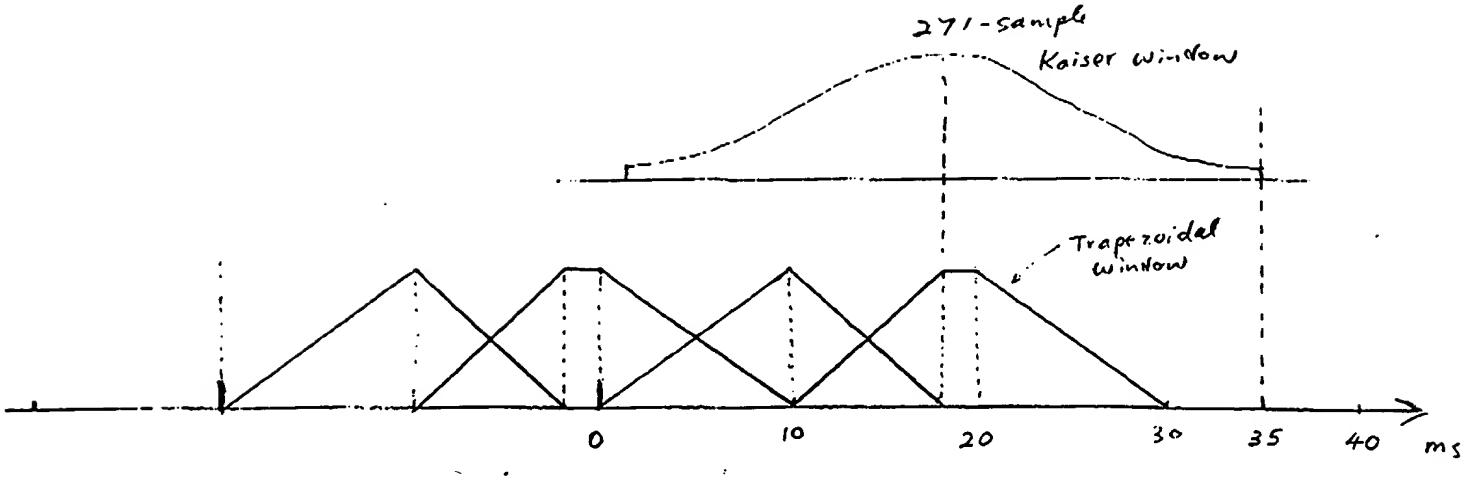

271-sample
Kaiser window
Trapezoidal
window
0
10
20
30
35
40
ms

Fig. 8. A

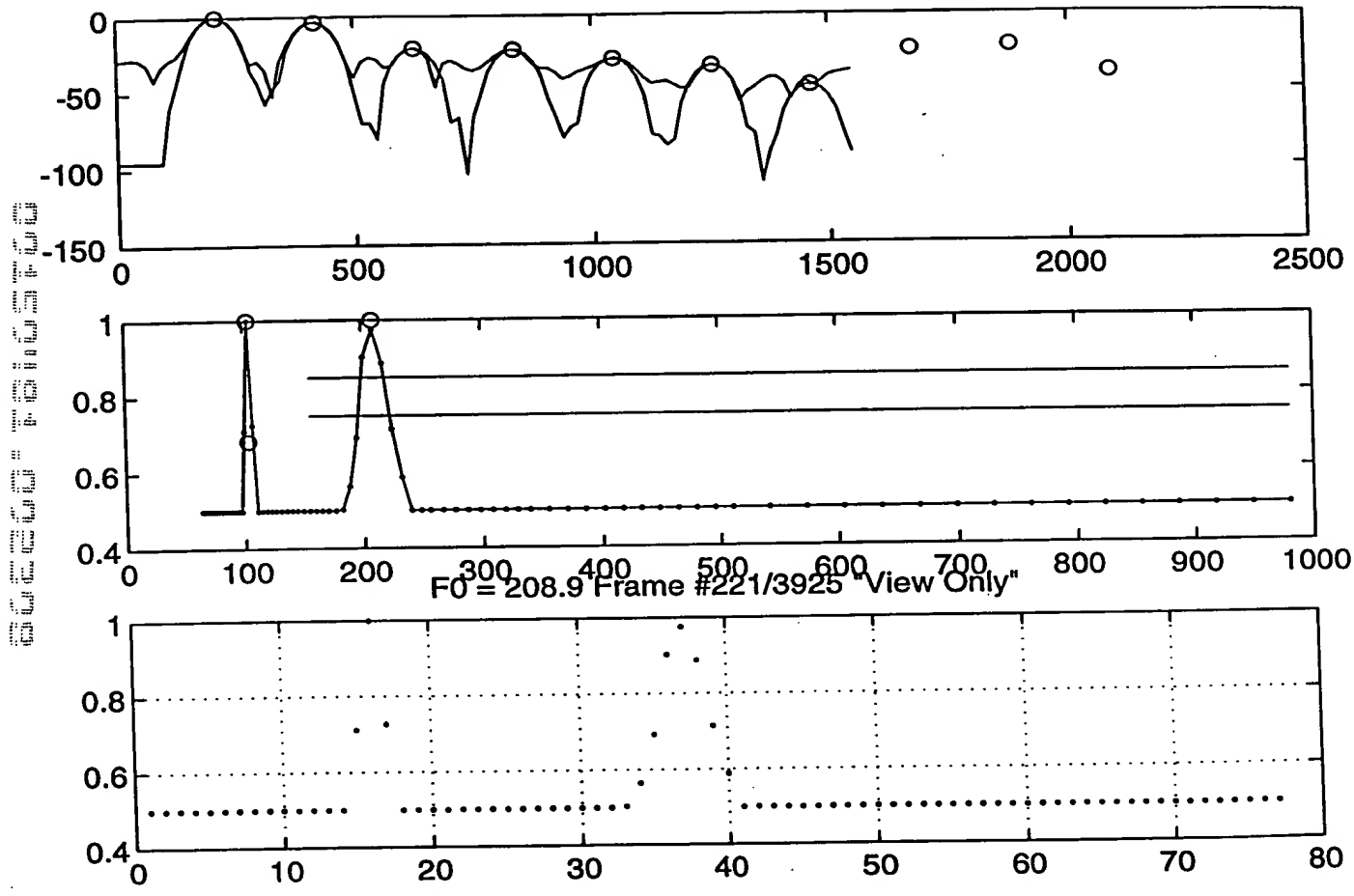

0
-50
-100
-150
0
500
1000
1500
2000
2500
1
0.8
0.6
0.4
0
100
200
300
400
500
600
700
800
900
1000
F0 = 208.9 Frame #221/3925 "View Only"
1
0.8
0.6
0.4
0
10
20
30
40
50
60
70
80

Fig. 9A

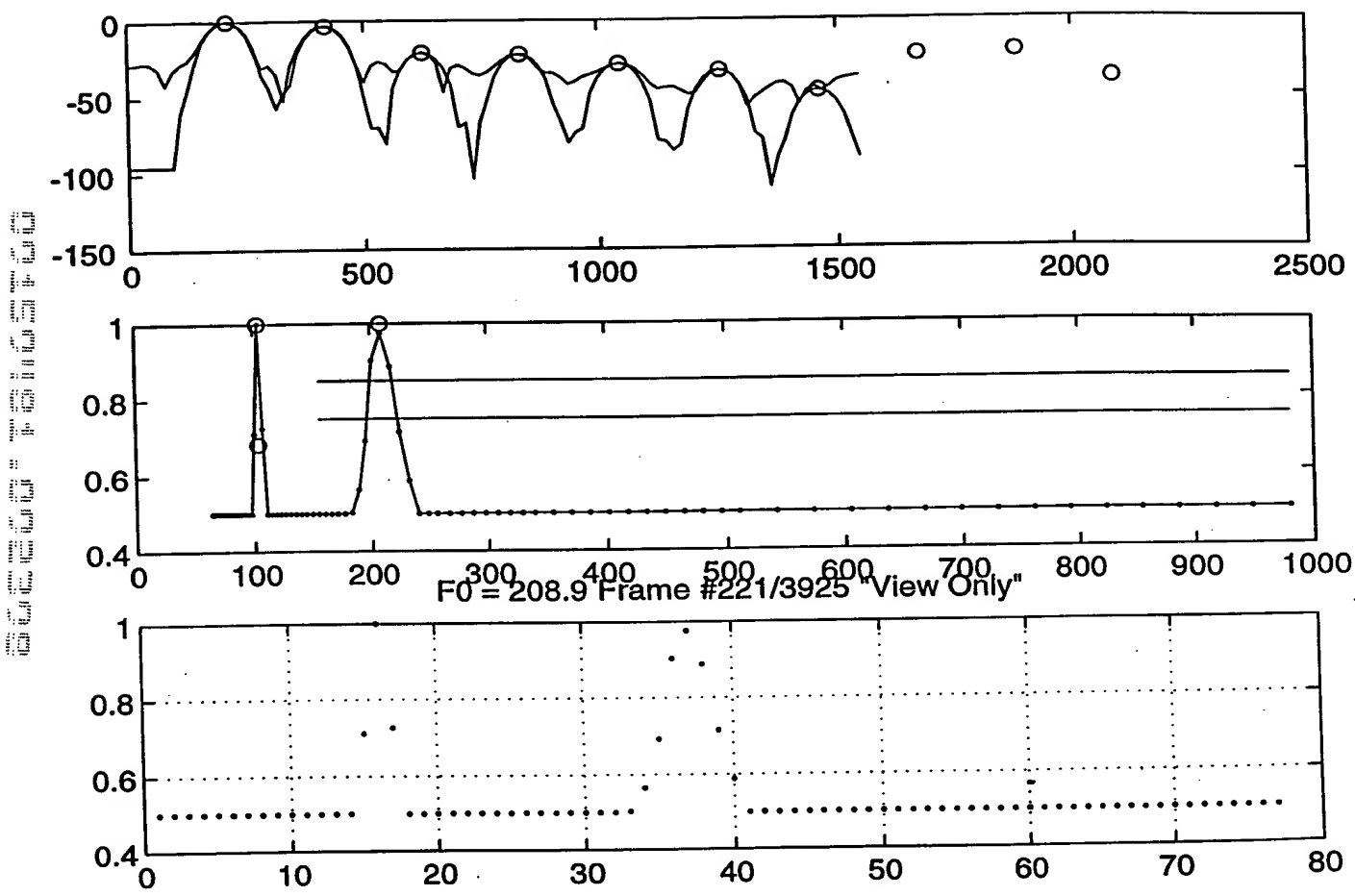

0
-50
-100
-150
0
500
1000
1500
2000
2500
1
0.8
0.6
0.4
0
100
200
300
400
500
600
700
800
900
1000
F0 = 208.9 Frame #221/3925 "View Only"
1
0.8
0.6
0.4
0
10
20
30
40
50
60
70
80

Fig. 9B

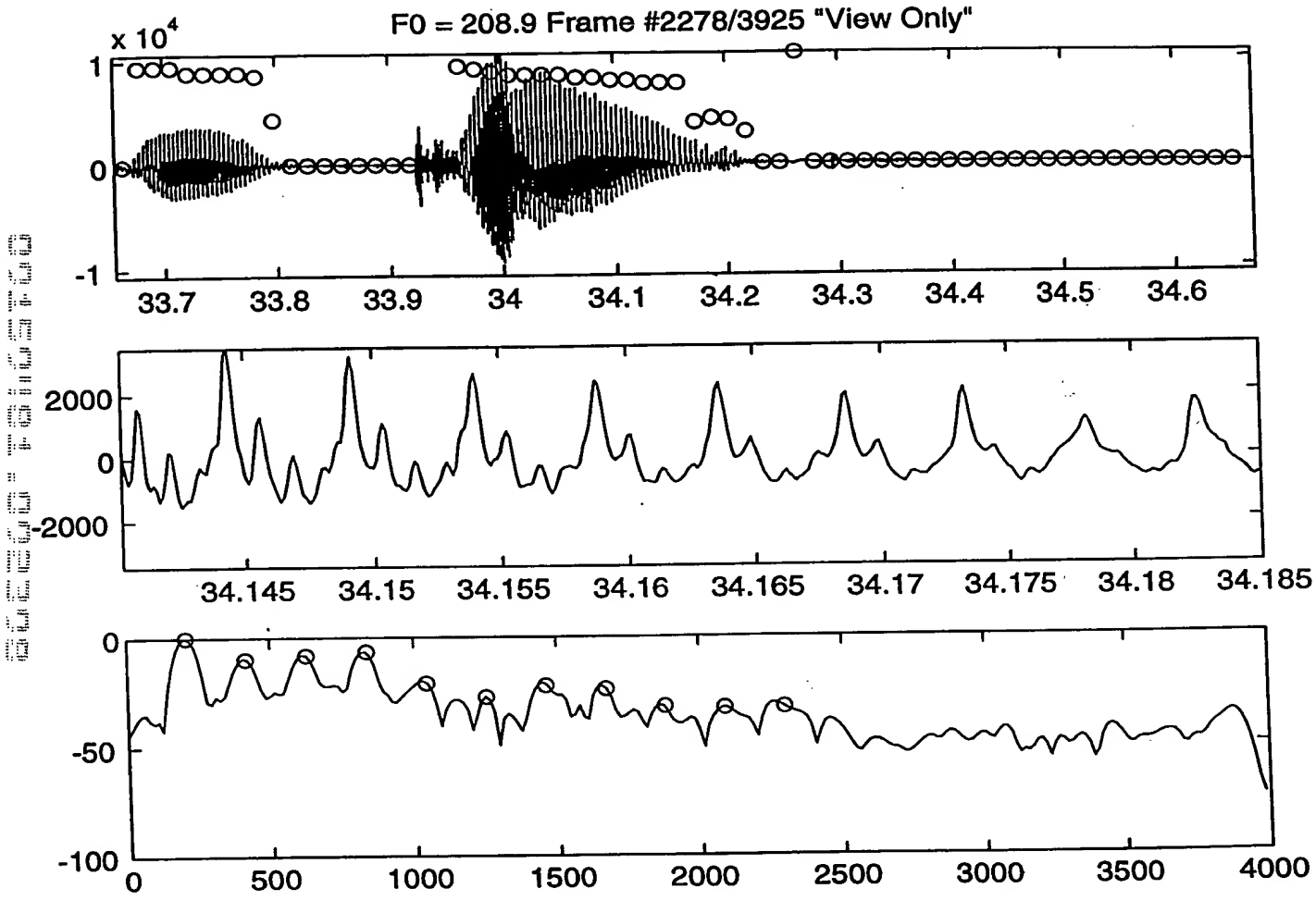

F0 = 208.9 Frame #2278/3925 "View Only"
x 10^4
1
0
-1
33.7
33.8
33.9
34
34.1
34.2
34.3
34.4
34.5
34.6
2000
0
-2000
34.145
34.15
34.155
34.16
34.165
34.17
34.175
34.18
34.185
0
-50
-100
0
500
1000
1500
2000
2500
3000
3500
4000

Fig. 9C

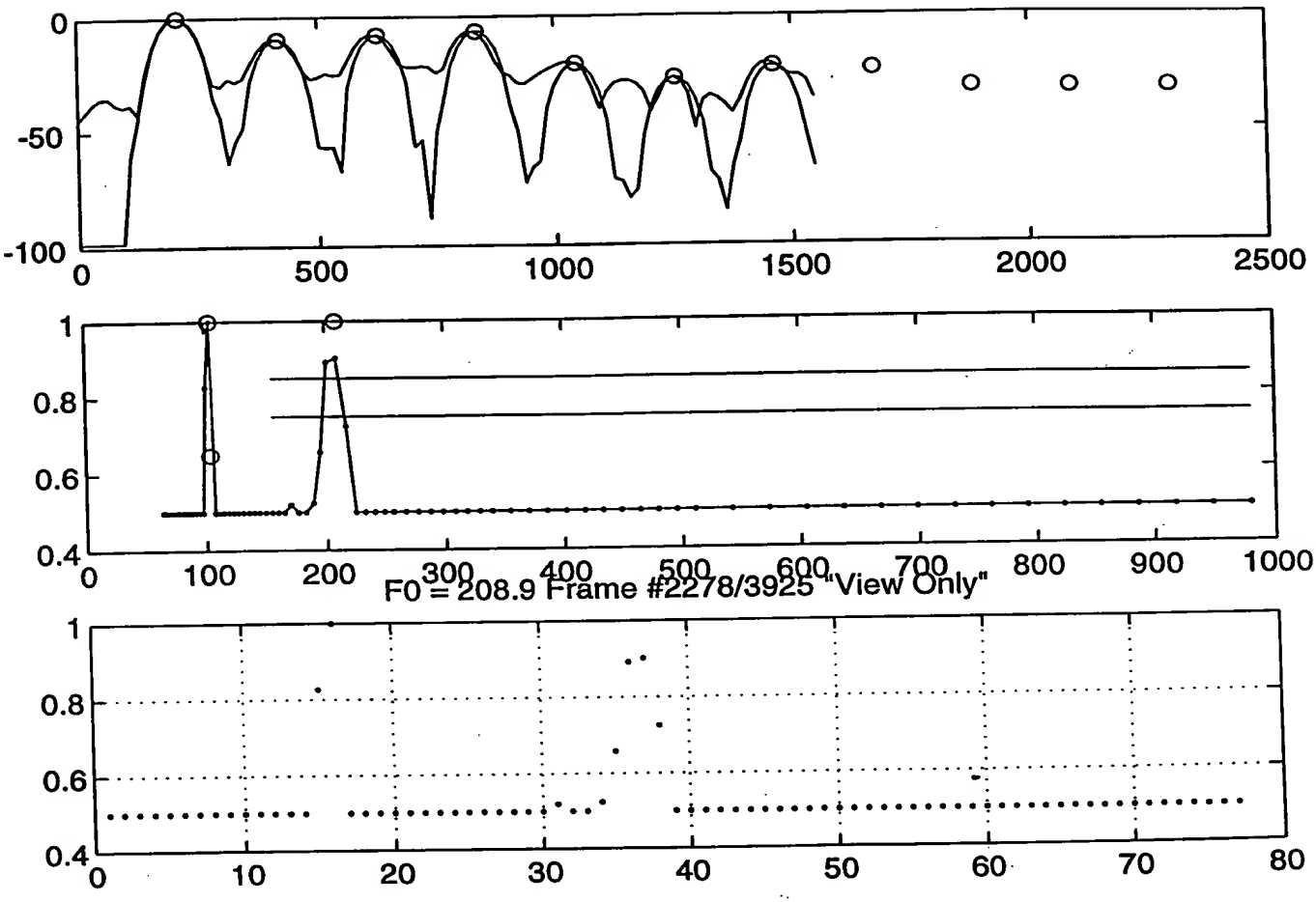

0
-50
-100
0
500
1000
1500
2000
2500
1
0.8
0.6
0.4
0
100
200
300
400
500
600
700
800
900
1000
F0 = 208.9 Frame #2278/3925 "View Only"
1
0.8
0.6
0.4
0
10
20
30
40
50
60
70
80

Fig. 9D

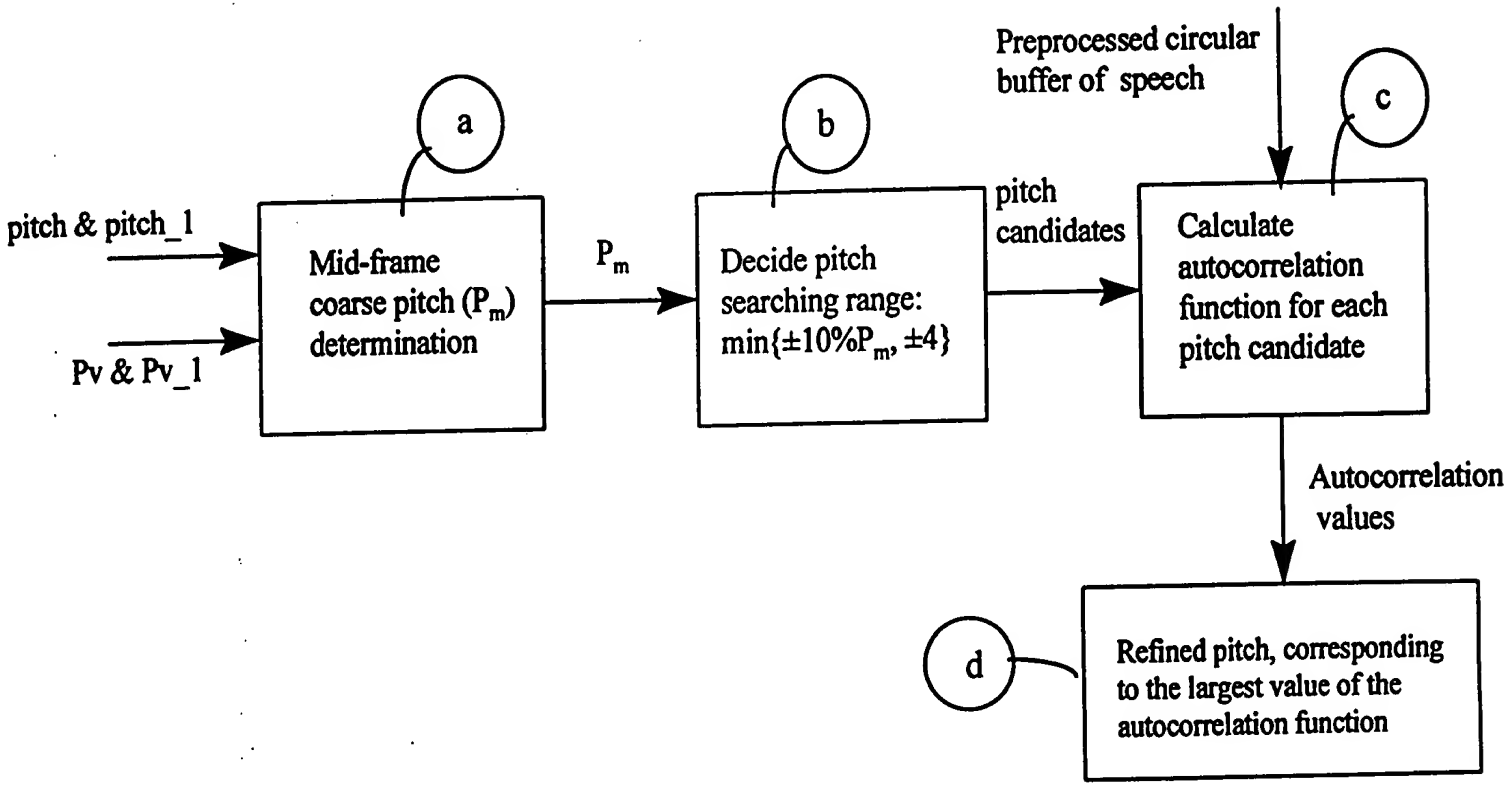

Preprocessed circular buffer of speech
a
b
c
pitch & pitch_1
Pv & Pv_1
Mid-frame coarse pitch (P_m) determination
P_m
Decide pitch searching range: min{±10%P_m, ±4}
pitch candidates
Calculate autocorrelation function for each pitch candidate
Autocorrelation values
d
Refined pitch, corresponding to the largest value of the autocorrelation function

Fig. 10

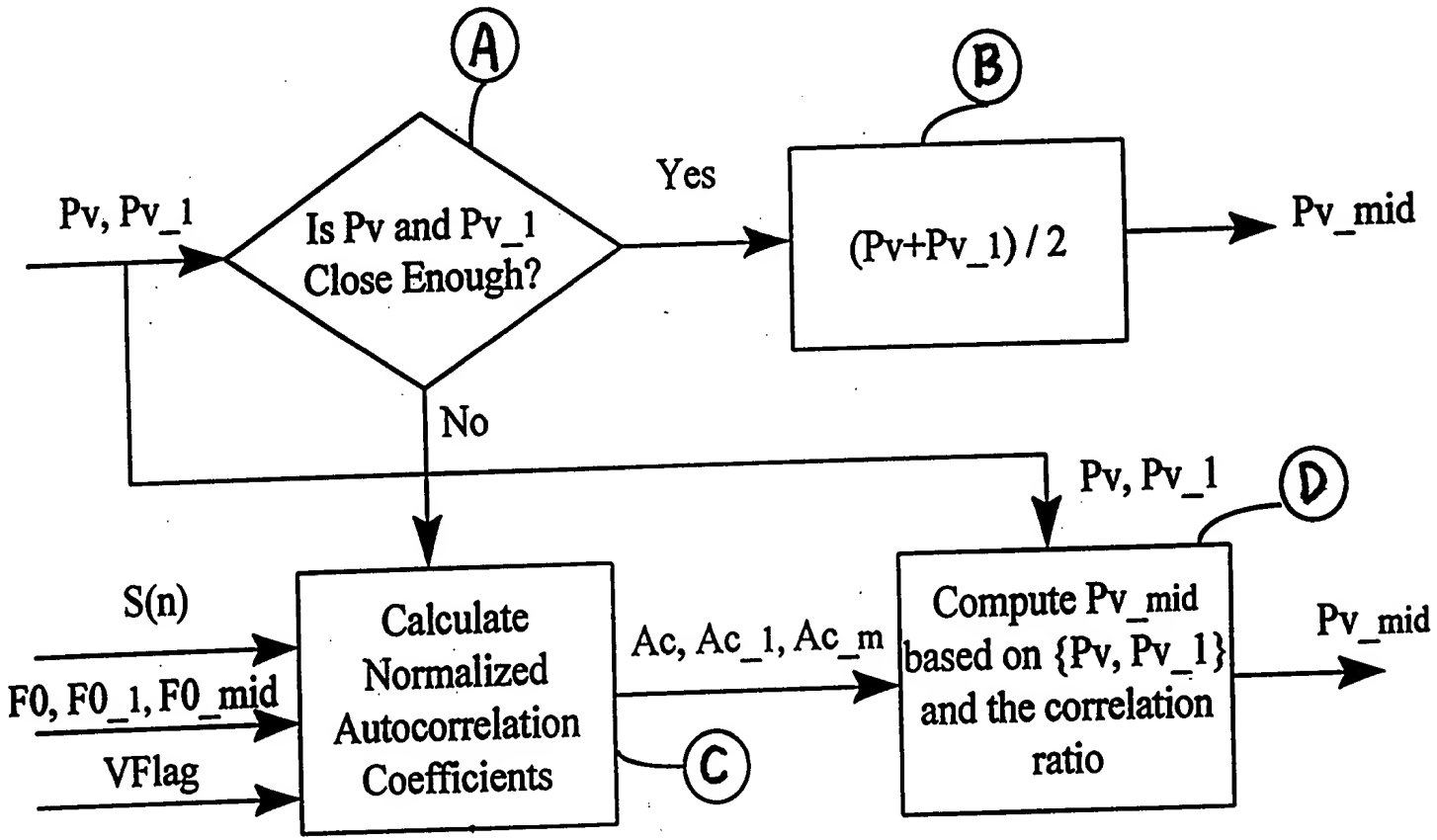


Fig. 11

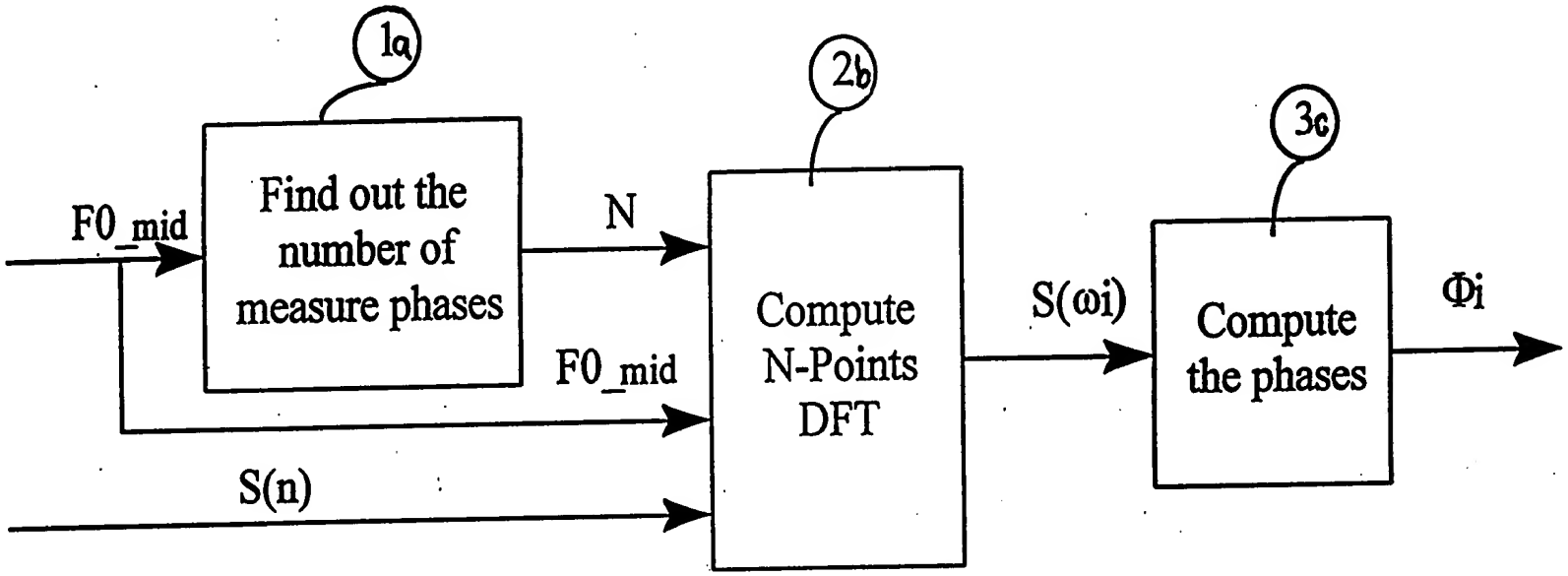

1a
2b
3c
F0_mid
Find out the number of measure phases
N
F0_mid
S(n)
Compute N-Points DFT
S(ωi)
Compute the phases
Φi

Fig. 12

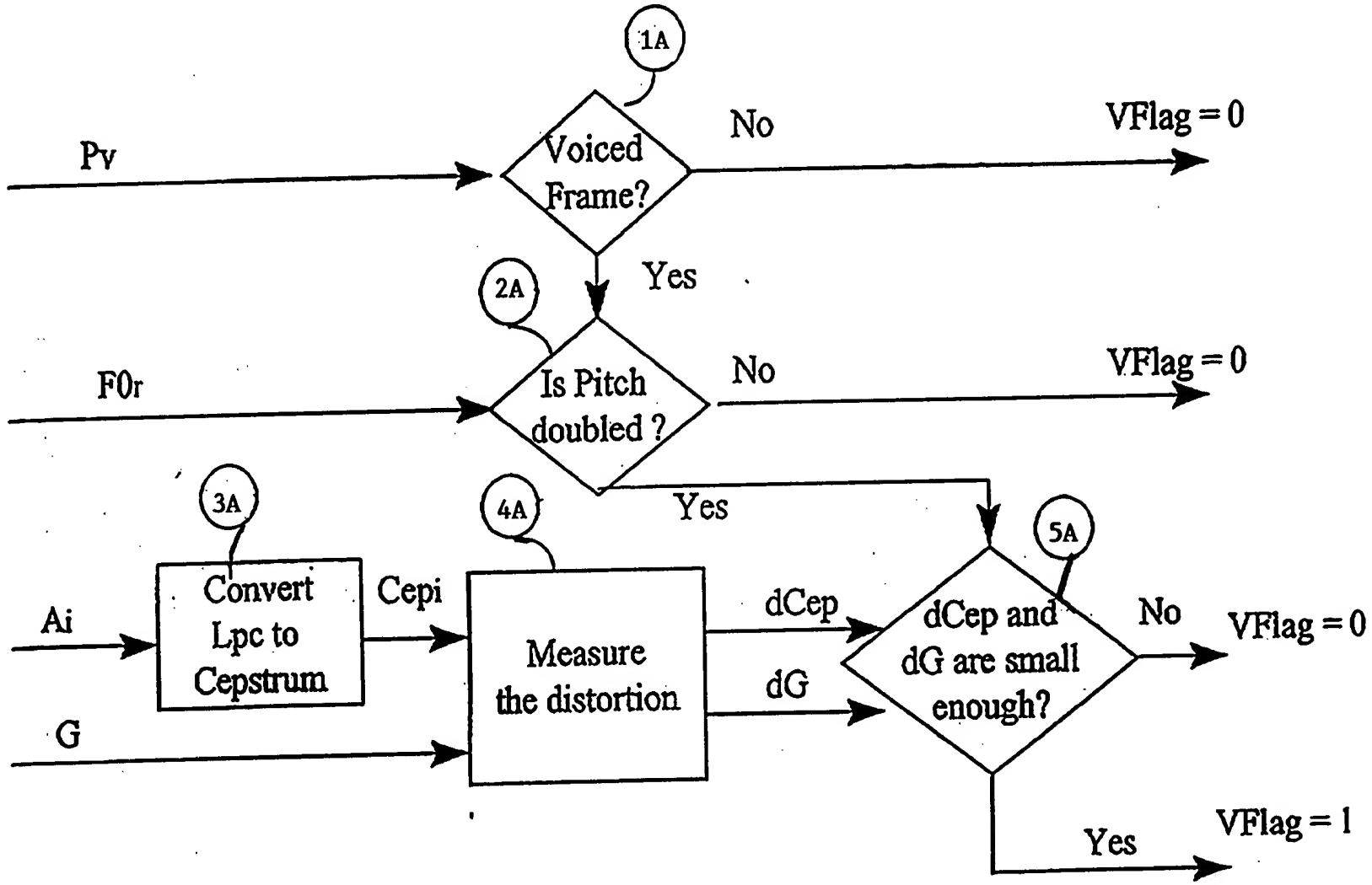

1A
Pv
Voiced Frame?
No
VFlag = 0
Yes
2A
F0r
Is Pitch doubled ?
No
VFlag = 0
Yes
3A
4A
5A
Ai
Convert Lpc to Cepstrum
Cepi
G
Measure the distortion
dCep
dG
dCep and dG are small enough?
No
VFlag = 0
Yes
VFlag = 1

Fig. 13

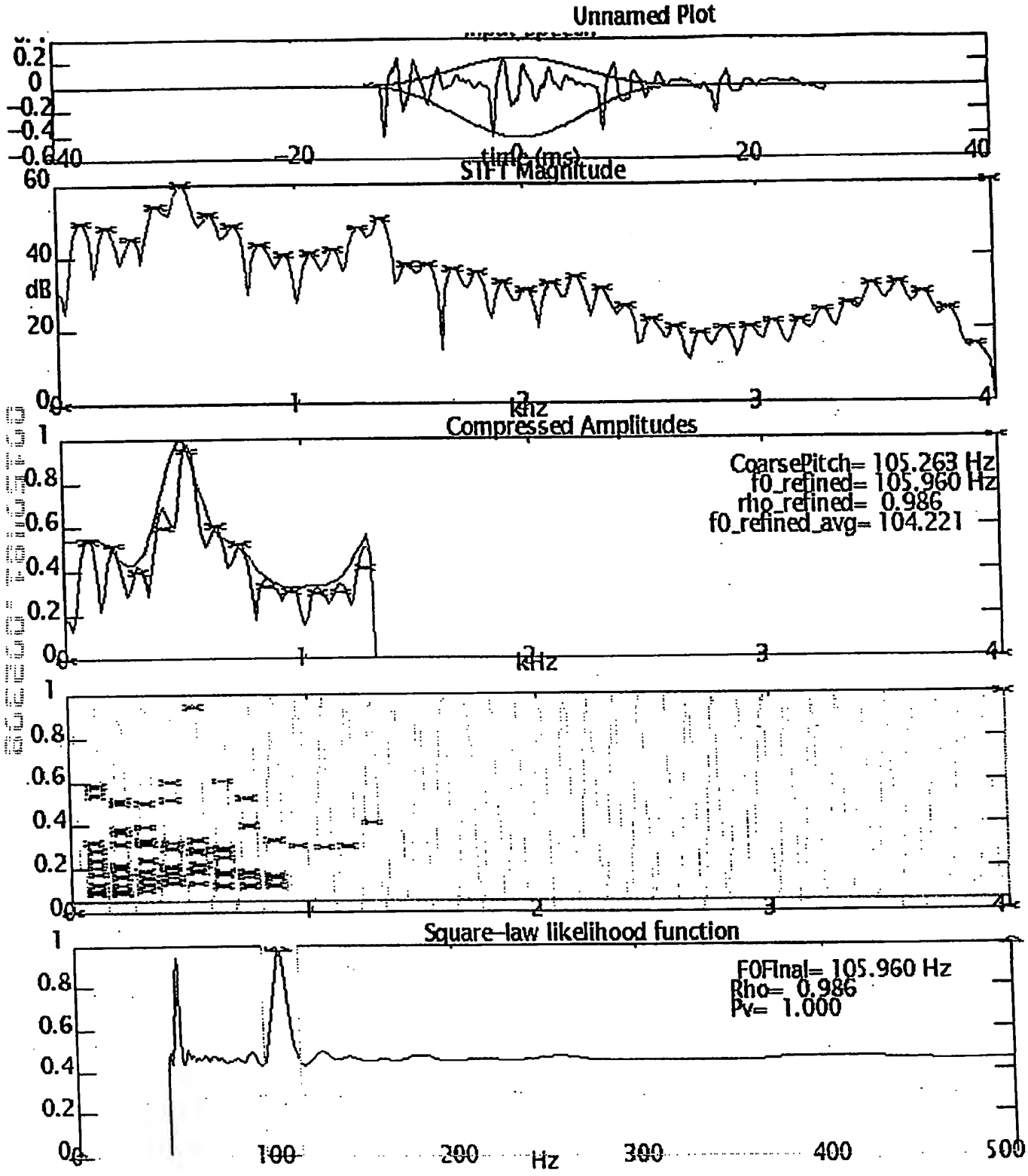

Unnamed Plot
0.2
0
-0.2
-0.4
-0.6
-40
-20
0
20
40
time (ms)
STFT Magnitude
60
40
dB
20
0
1
2
3
4
kHz
Compressed Amplitudes
CoarsePitch= 105.263 Hz
f0_refined= 105.960 Hz
rho_refined= 0.986
f0_refined_avg= 104.221
1
0.8
0.6
0.4
0.2
0
kHz
Square-law likelihood function
F0Final= 105.960 Hz
Rho= 0.986
Pv= 1.000
100
200
300
400
500
Hz

Fig. 14

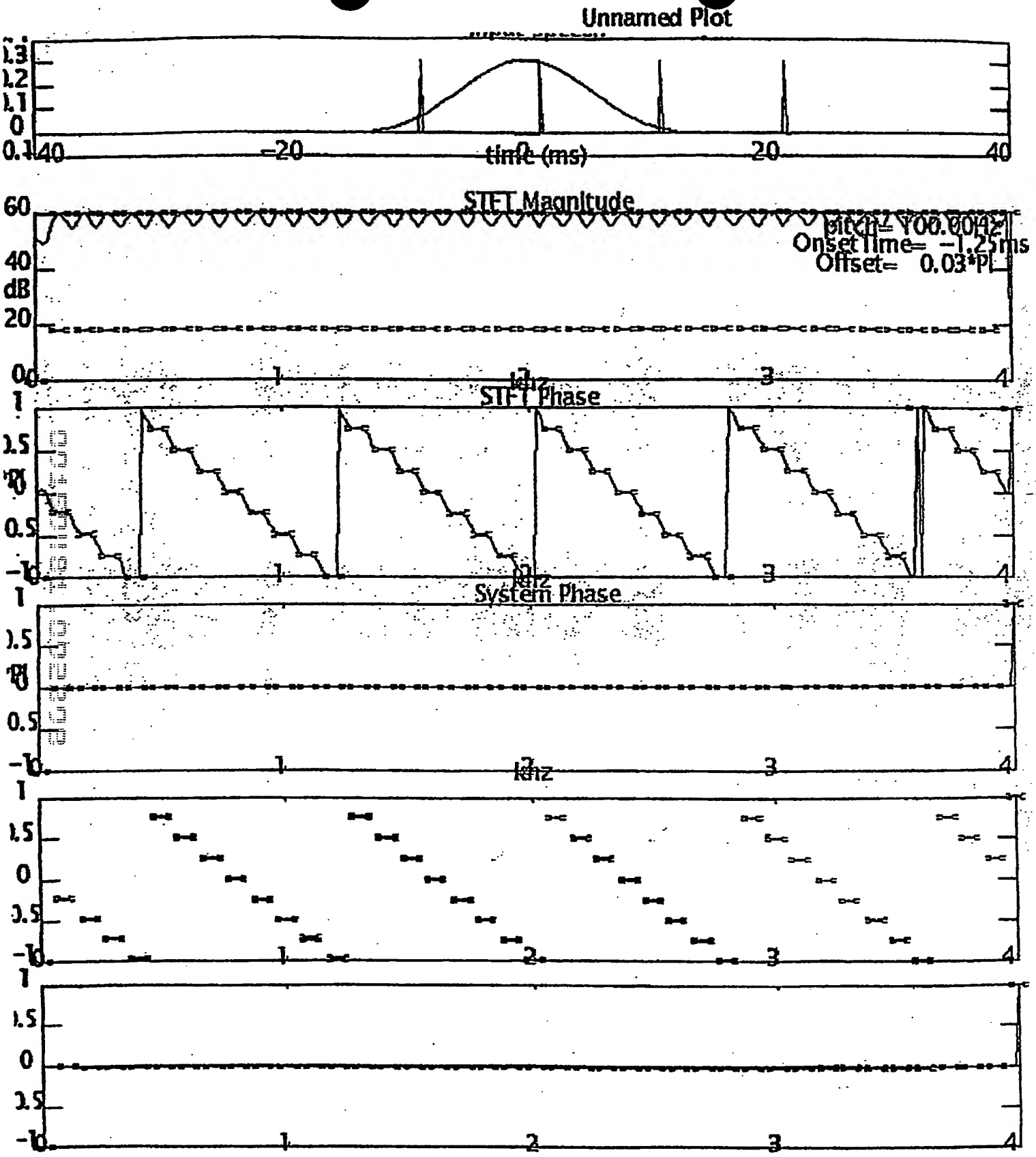

Unnamed Plot
1.3
1.2
1.1
0
0.1
-40
-20
0
20
40
time (ms)
STFT Magnitude
60
40
dB
20
0
1
2
3
4
kHz
pitch= 100.00Hz
Onset time= -1.25ms
Offset= 0.03*Pi
STFT Phase
1
0.5
0
-0.5
-1
System Phase
Excitation Phase
Residual Phase

Fig. 15

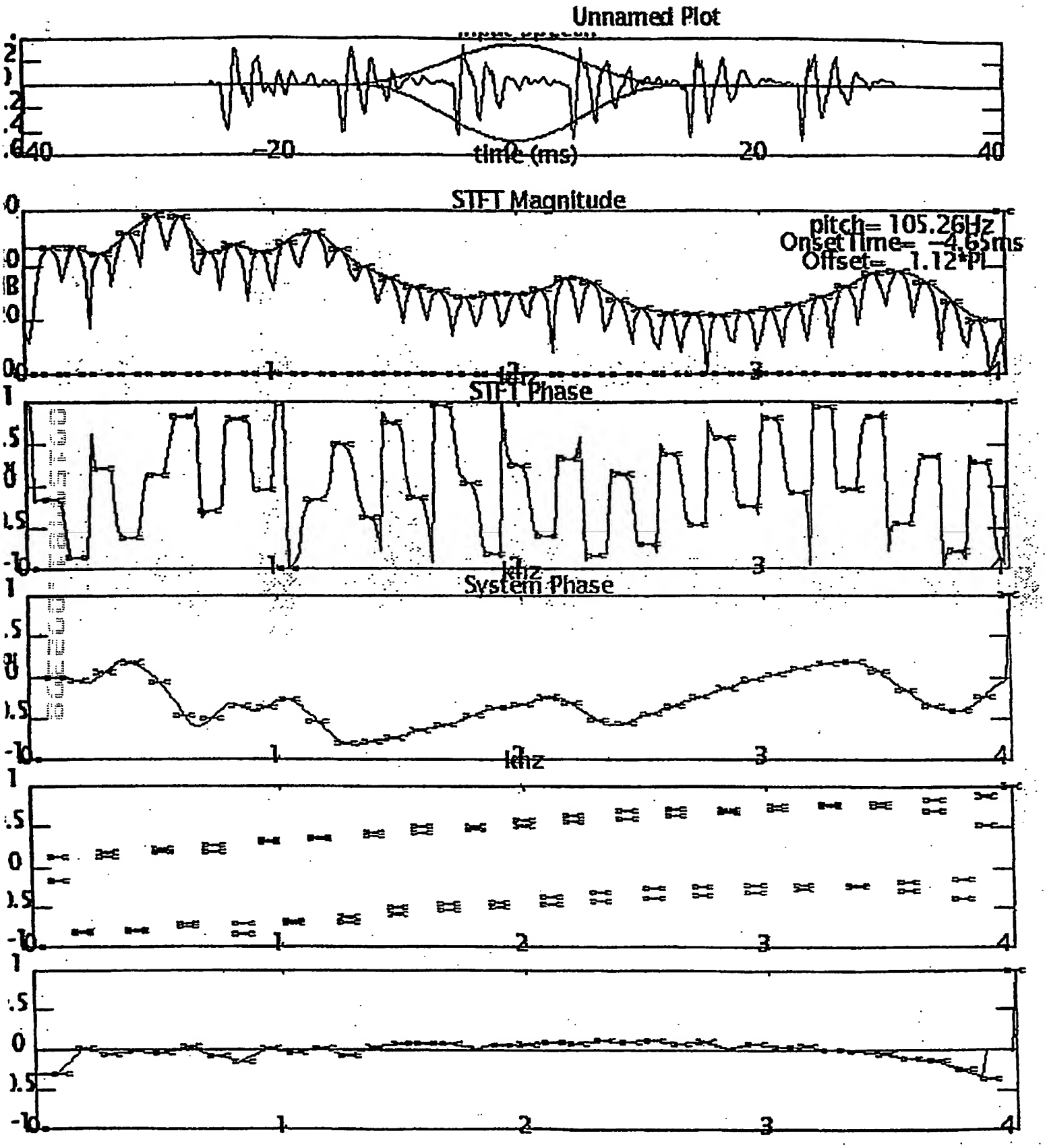

Unnamed Plot
time (ms)
STFT Magnitude
pitch= 105.26Hz
Onset Time= -4.65ms
Offset= 1.12*PL
STFT Phase
khz
System Phase
khz

Fig. 16

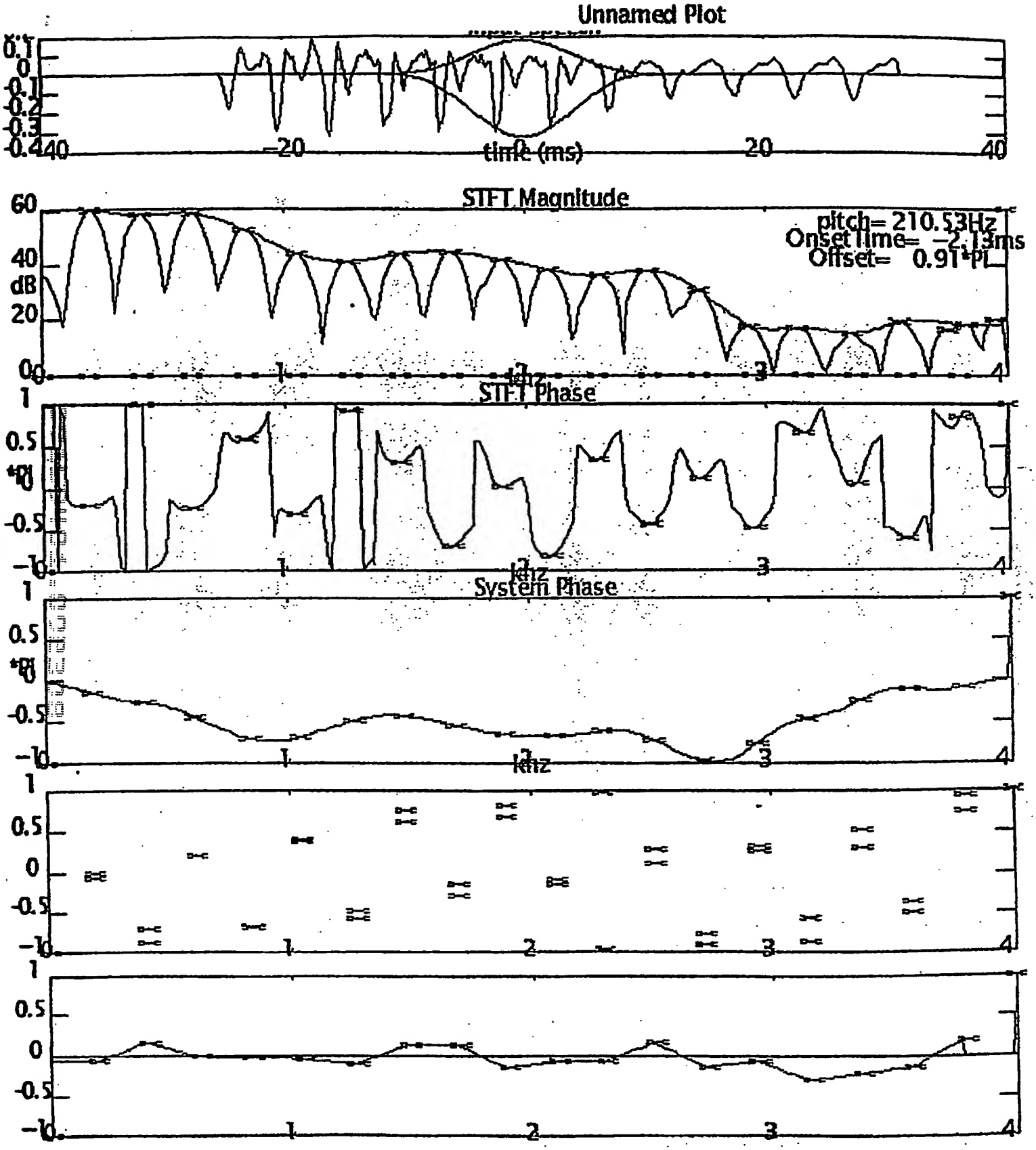

Unnamed Plot
0.1
0
-0.1
-0.2
-0.3
-0.4
-40
-20
time (ms)
20
40
STFT Magnitude
pitch= 210.53Hz
Onset time= -2.13ms
Offset= 0.91*pi
60
40
dB
20
0
1
2
3
4
kHz
STFT Phase
1
0.5
*pi
0
-0.5
-1
System Phase

Fig. 17

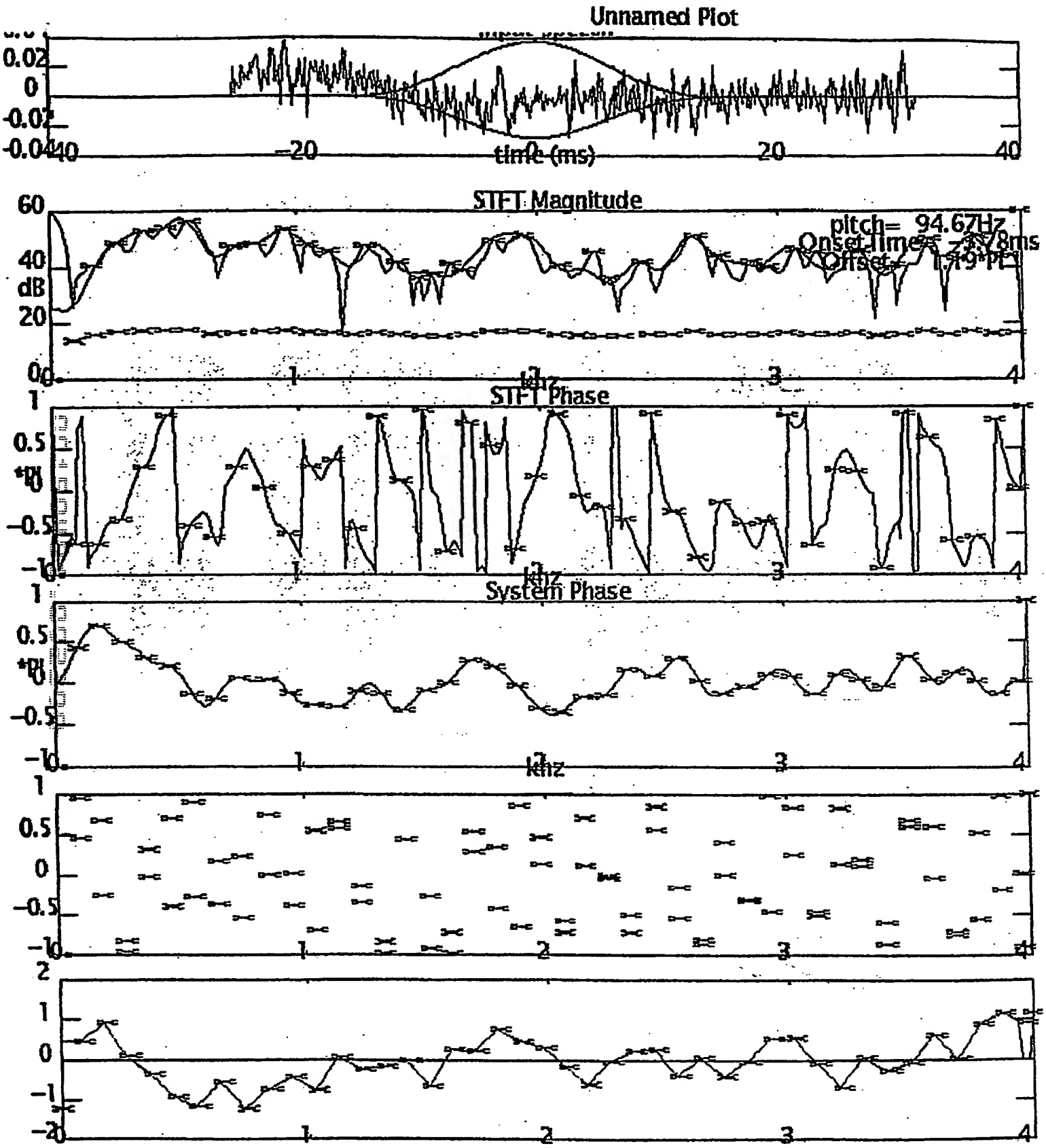

Unnamed Plot
0.02
0
-0.02
-0.04
-40
-20
time (ms)
20
40
STFT Magnitude
pitch= 94.67Hz
60
40
dB
20
0
1
khz
3
4
STFT Phase
1
0.5
0
-0.5
-1
System Phase
2
-2

Fig. 18

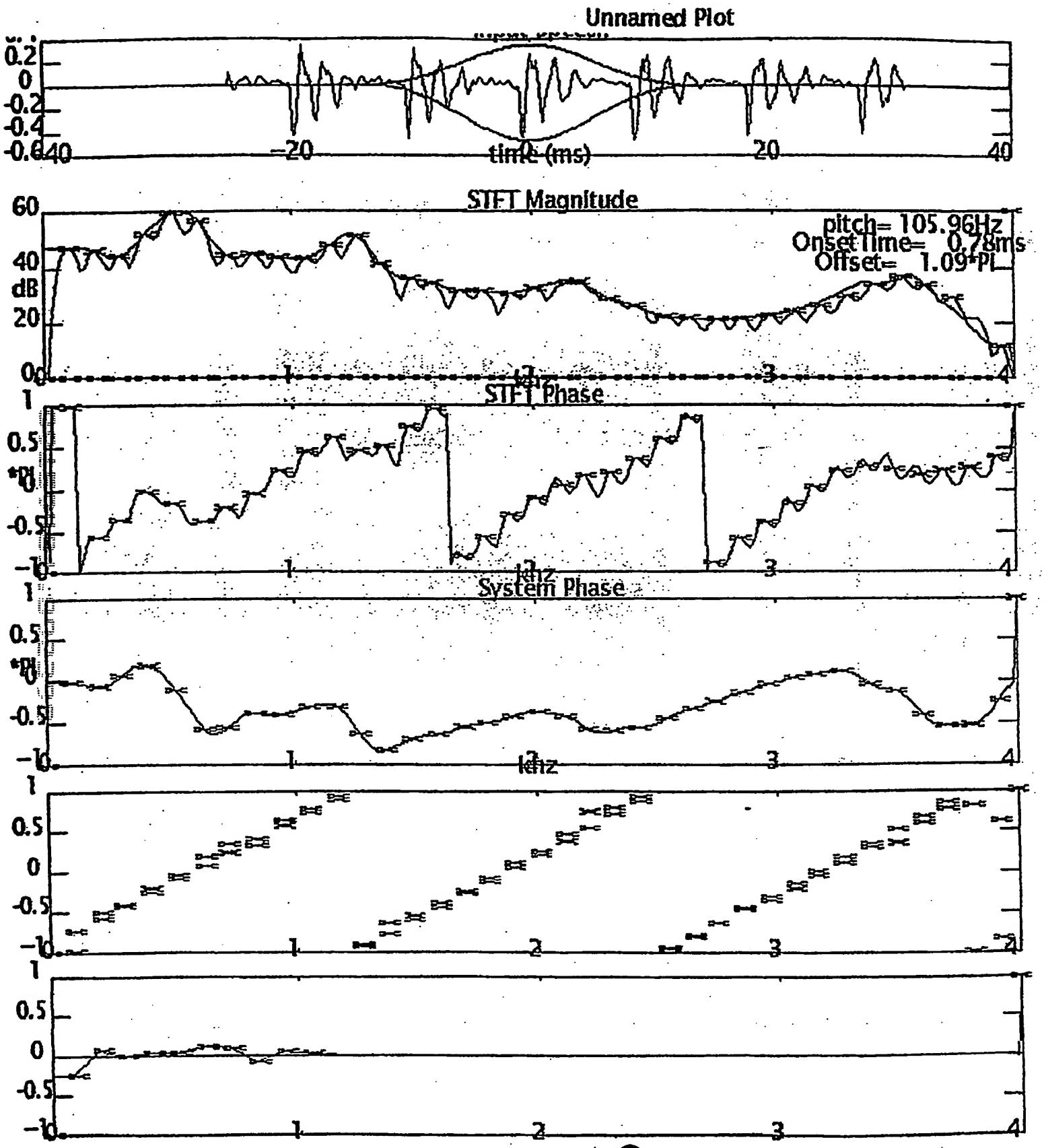
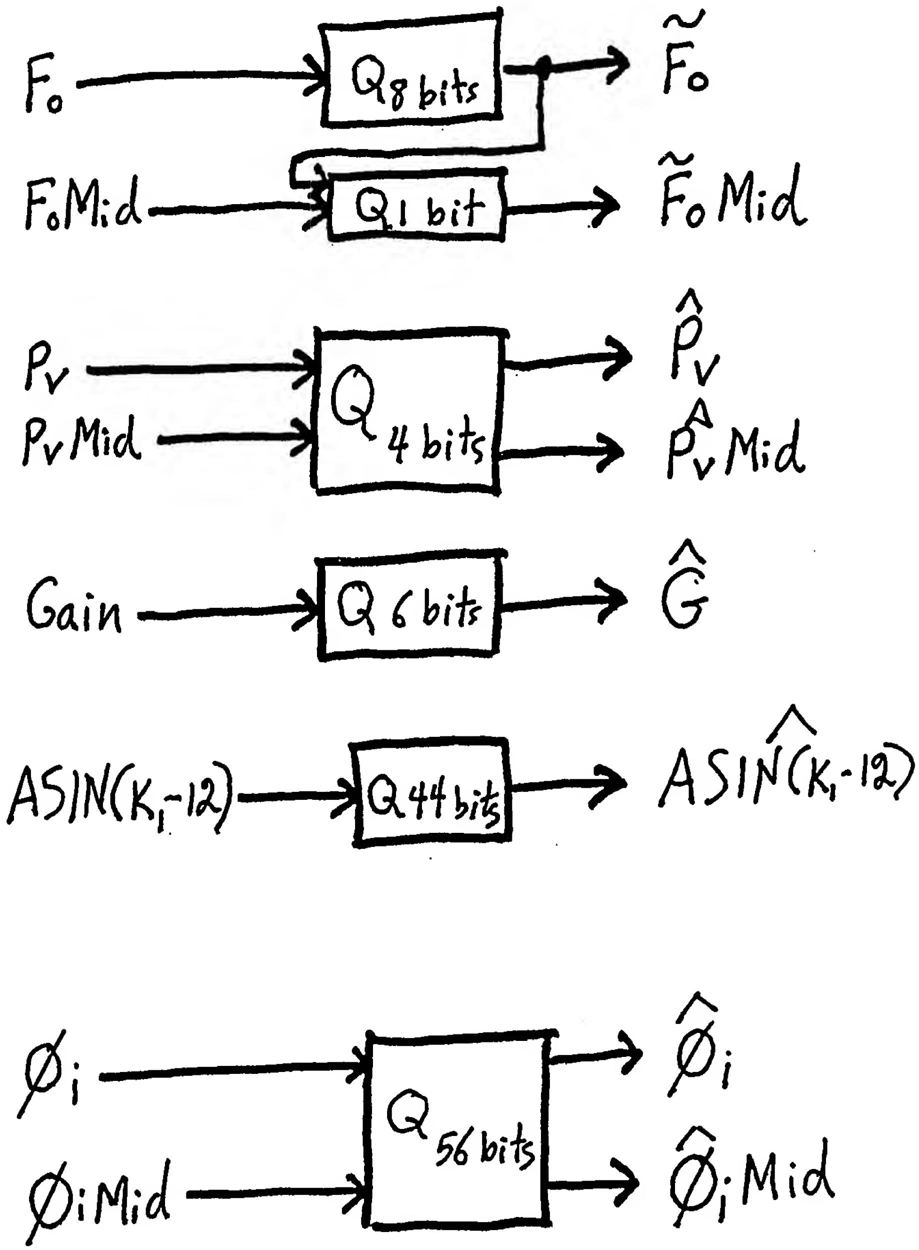

Unnamed Plot
0.2
0
-0.2
-0.4
-0.6
-40
-20
time (ms)
20
40
STFT Magnitude
60
40
dB
20
0
pitch= 105.96Hz
Onset time= 0.78ms
Offset= 1.09*PI
1
2
3
4
kHz
STFT Phase
1
0.5
*PI
0
-0.5
-1
System Phase

Fig. 19



F_0
F_0 Mid
Q 8 bits
Q 1 bit
$\tilde{F}_0$
$\tilde{F}_0$ Mid
P_V
P_V Mid
Q 4 bits
$\hat{P}_V$
$\hat{P}_V$ Mid
Gain
Q 6 bits
$\hat{G}$
$ASIN(K_i-12)$
Q 44 bits
$\widehat{ASIN(K_i-12)}$
ϕ_i
ϕ_i Mid
Q 56 bits
$\hat{\phi}_i$
$\hat{\phi}_i$ Mid

Fig. 20

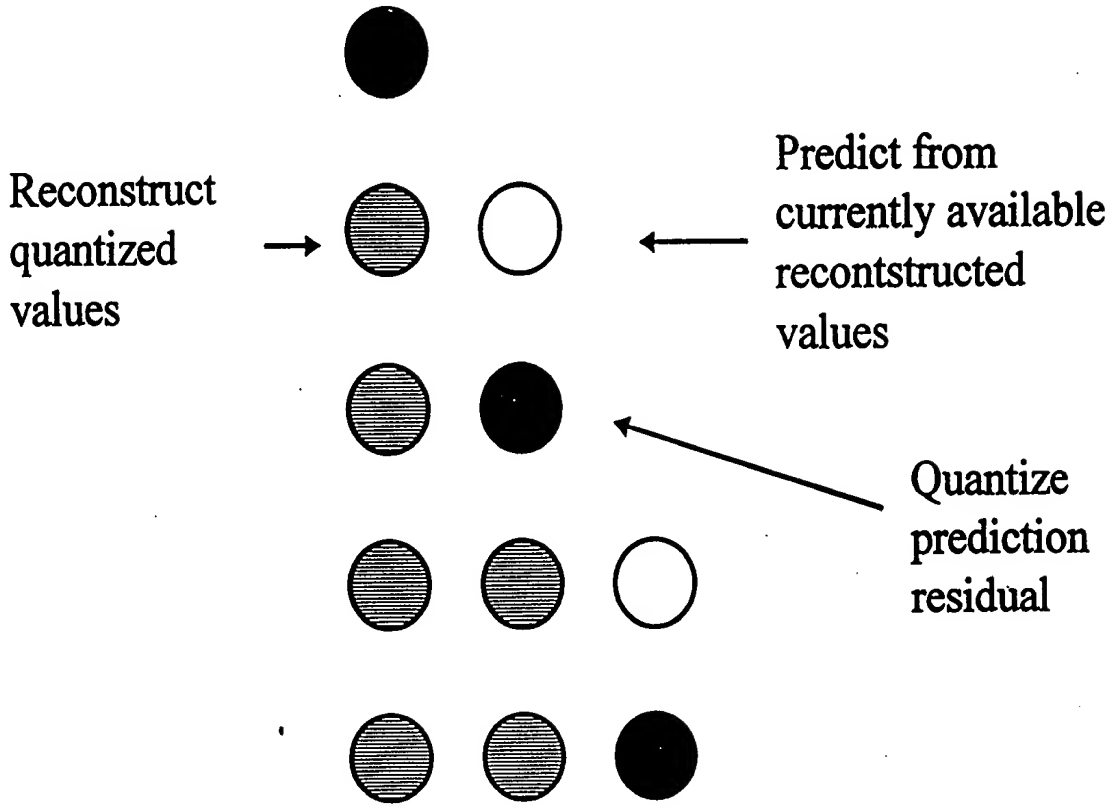


Fig. 21

Spectral Quantization

prediction of $AS(i)$

$$\hat{AS}(i) = \sum_{j=1}^{i-1} a_{ij}\, \tilde{AS}(j)$$

$$\begin{bmatrix} \hat{AS}(1) \\ \hat{AS}(2) \\ \vdots \\ \hat{AS}(12) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ a_{21} & 0 & & & \\ a_{31} & a_{32} & 0 & & \\ \vdots & & & & \\ a_{12,1} & a_{12,2} & \cdots & & a_{12,11} \end{bmatrix} \begin{bmatrix} \tilde{AS}(1) \\ \tilde{AS}(2) \\ \vdots \\ \tilde{AS}(11) \end{bmatrix}$$

12×1 · 12×11 · 11×1

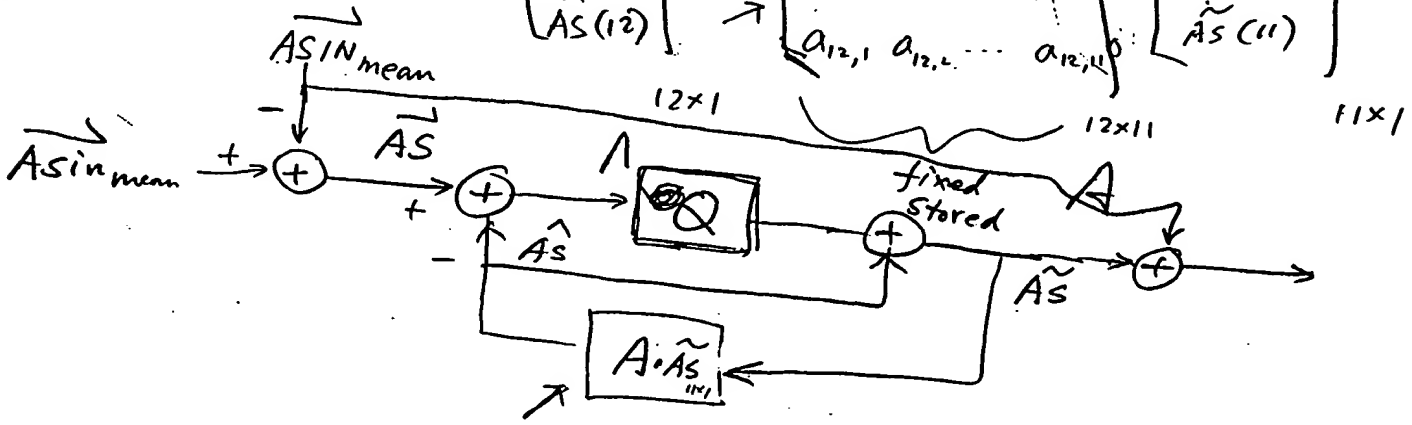
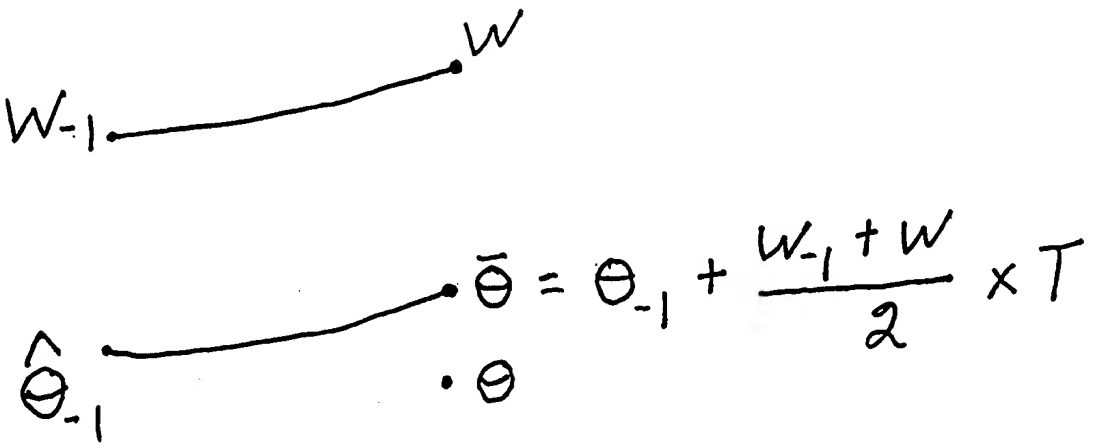


Fig. 21A

PHASE-PREDICTIVE CODING



PHASE RESIDUAL = $\theta - \bar{\theta}$

W_{-1} = frequency at previous frame
W = frequency at current frame
θ_{-1} = quantized phase at previous frame
$\bar{\theta}$ = predicted phase at current frame
θ = measured phase at current frame

Fig. 22A

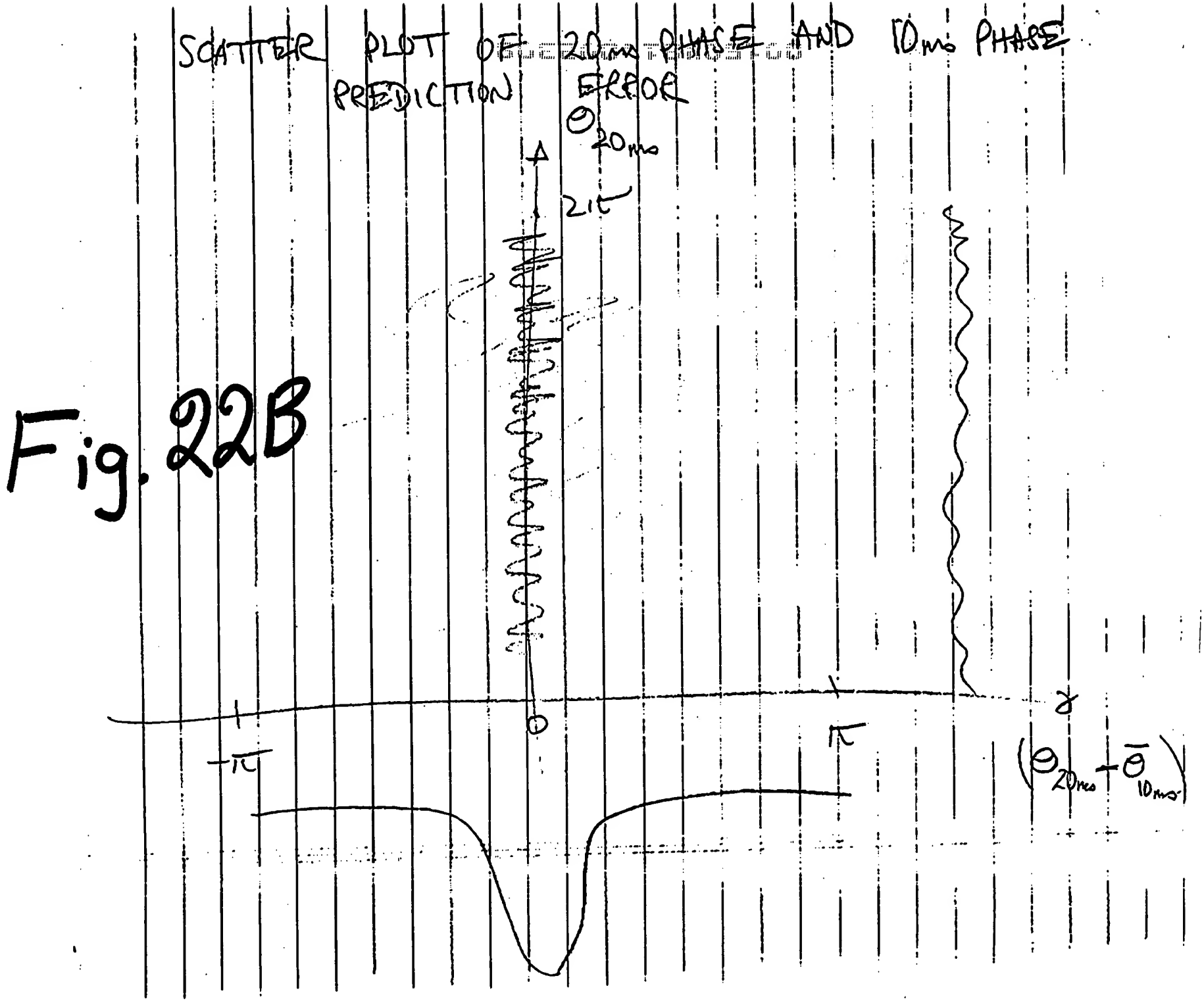
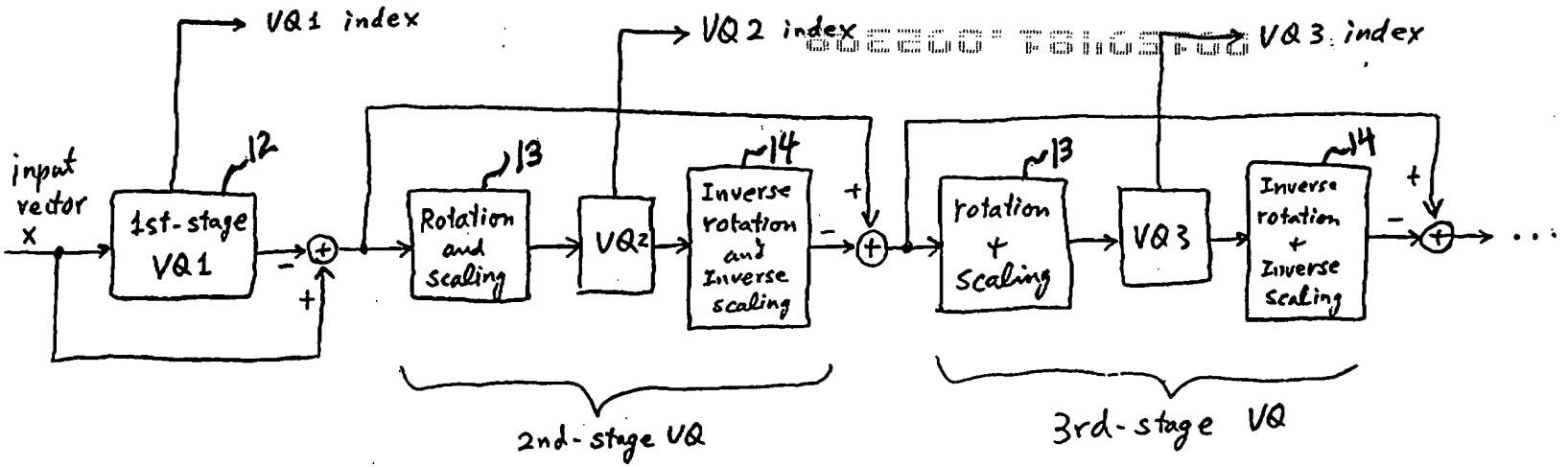
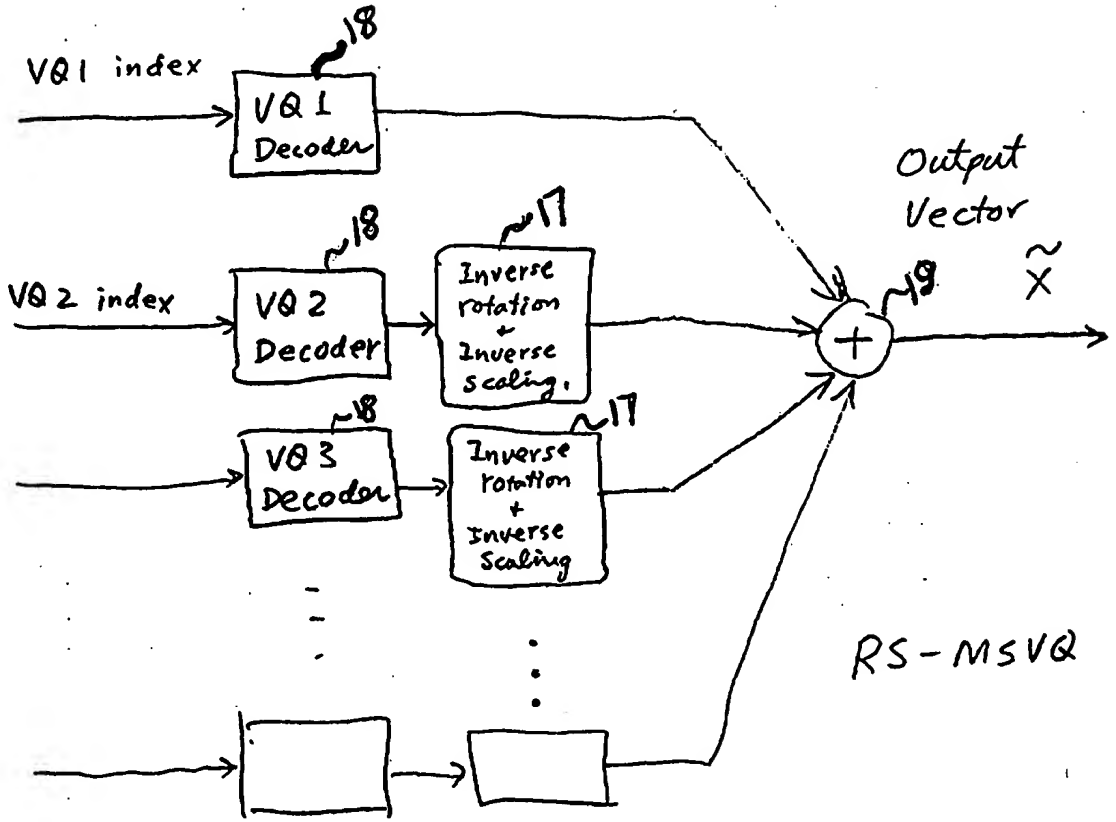

SCATTER PLOT OF 20ms PHASE AND 10ms PHASE
PREDICTION ERROR
θ_{20ms}
2π
$-\pi$
0
π
$(\theta_{20ms} + \bar{\theta}_{10ms})$

Fig. 22B



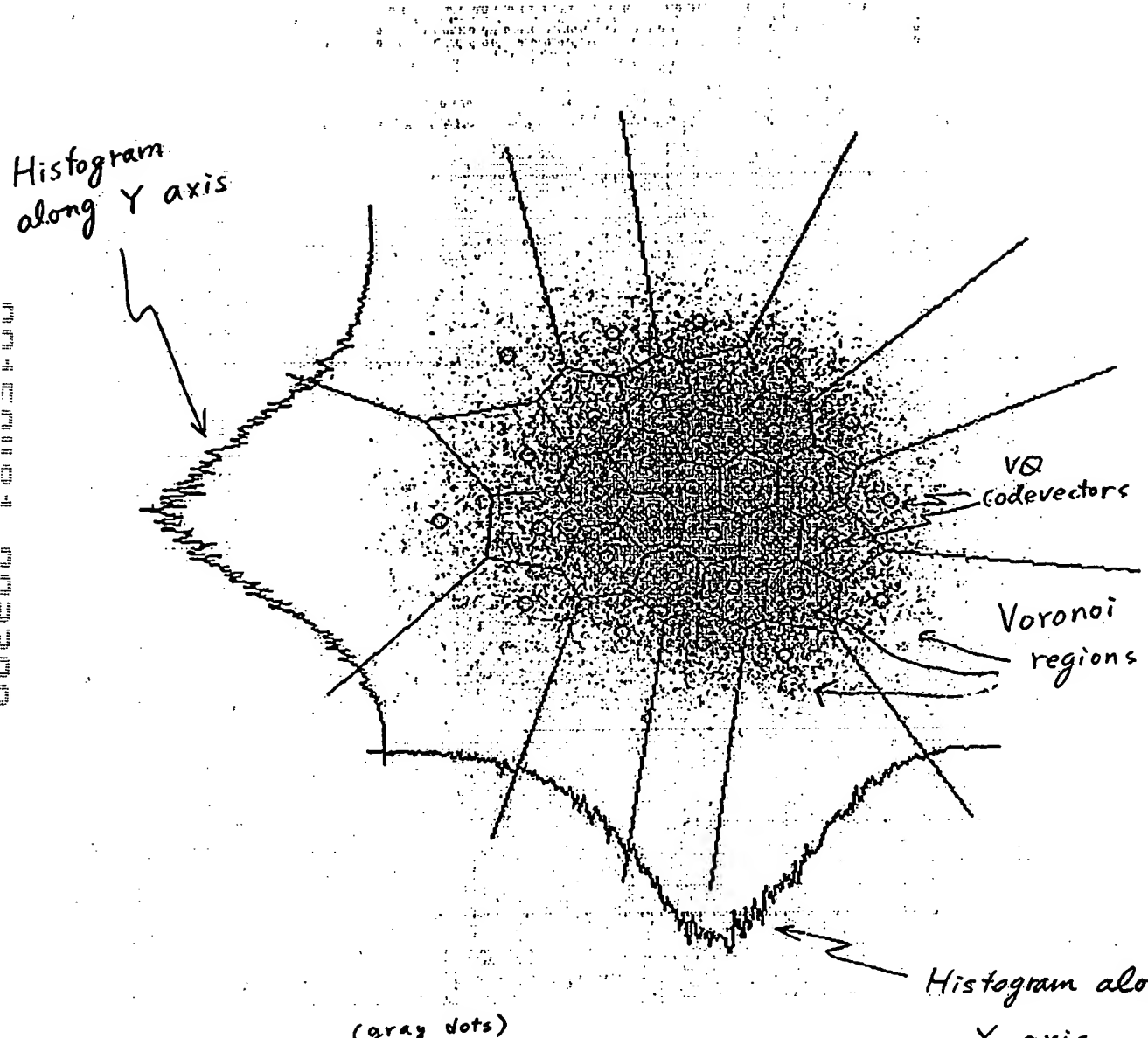
RS-MSVQ Encoder

Fig. 23A



RS-MSVQ Decoder

Fig. 23B



Scatter plot (gray dots) of 4th pair of ASIN(k) intra-frame prediction error, the histogram along each direction, and the corresponding 1st-stage 5-bit VQ codebook and Voronoi regions

Fig. 24A

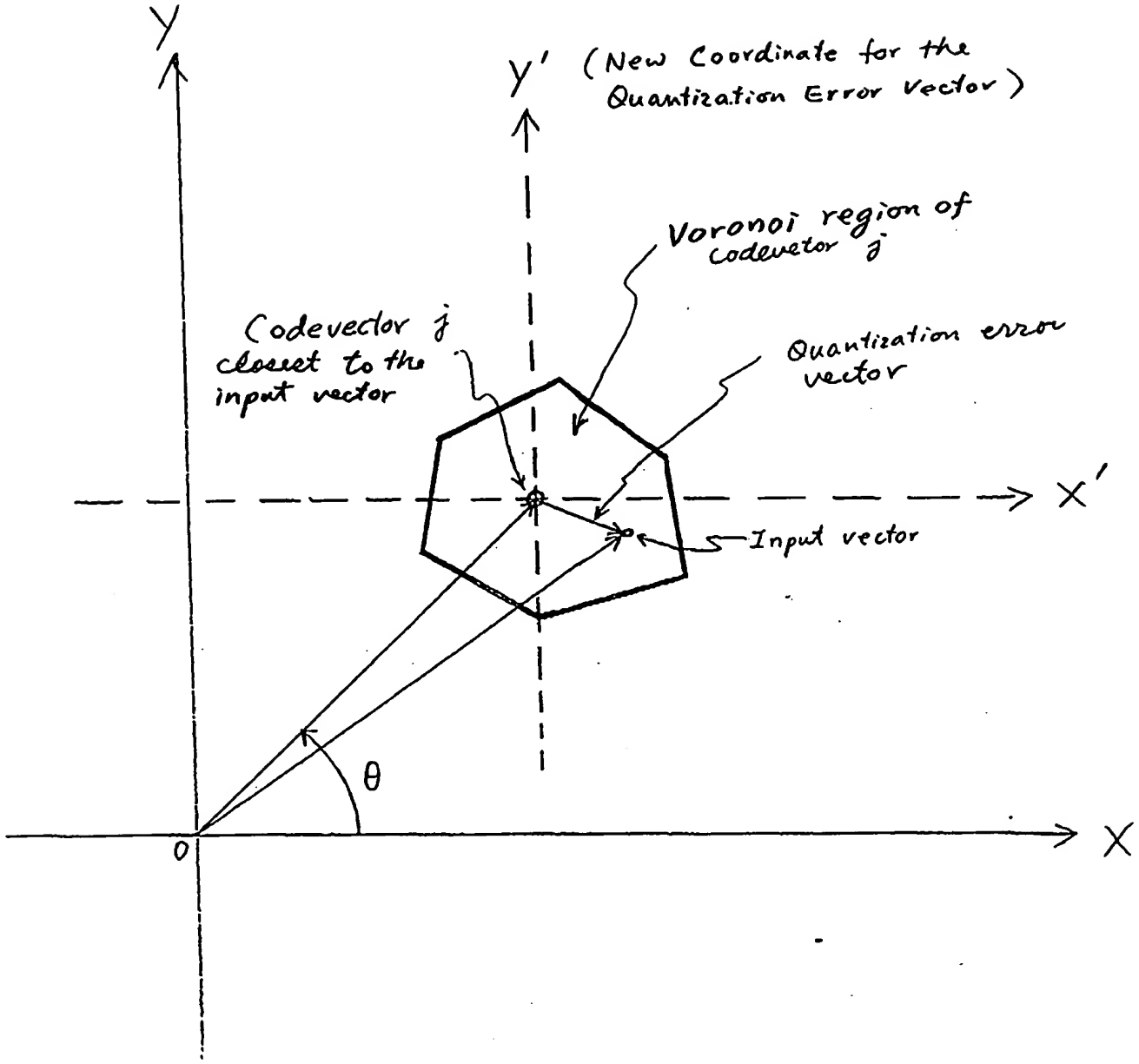
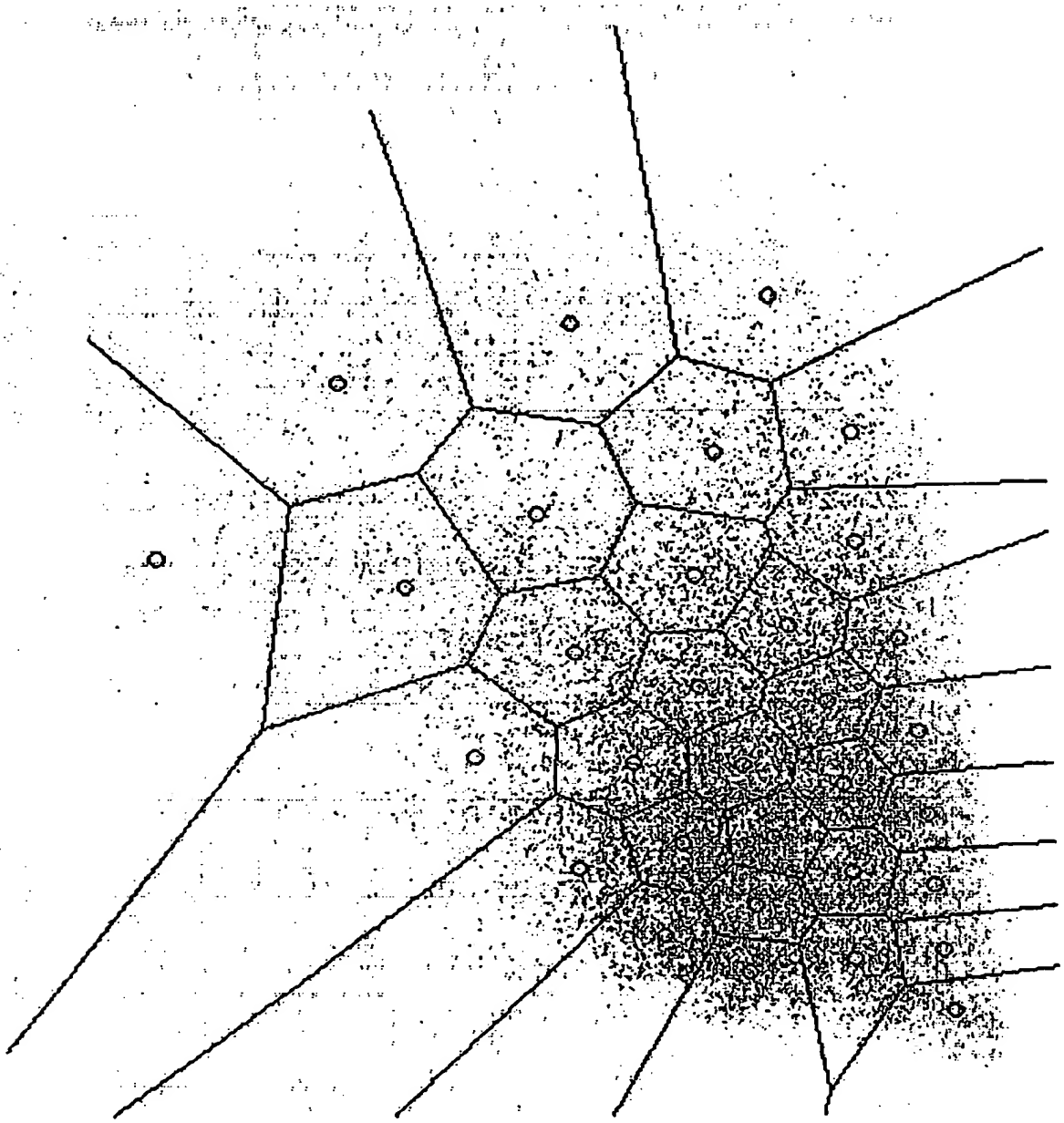


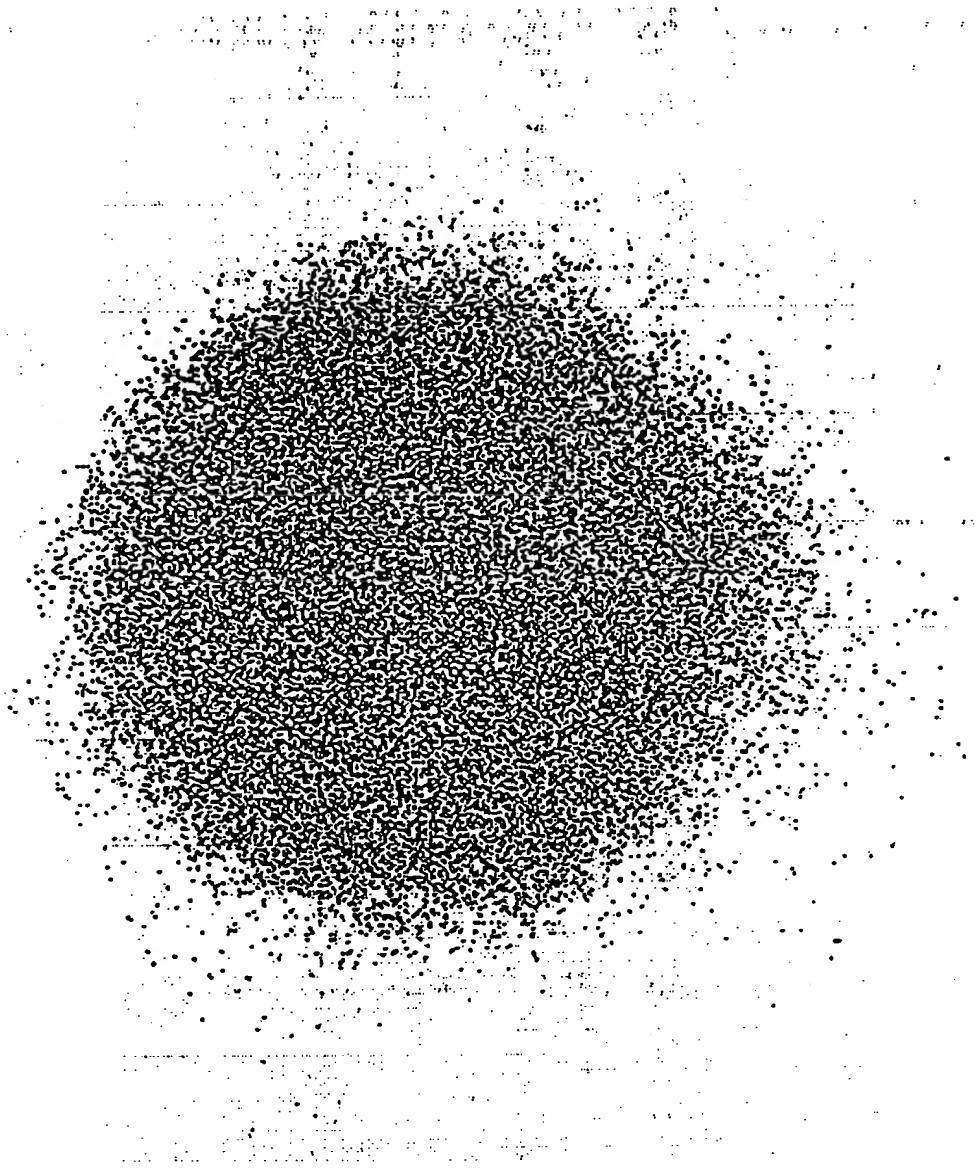
Fig. 24 B Illustration of the effect of subtracting the closest codevector for the input vector to get the quantization error vector.

Fig. 24B



Scatter plot of 1st pair of ASIN(k) (gray dots) and 1st-stage VQ codebook (small circles), and the corresponding Voronoi cells

Fig. 24C

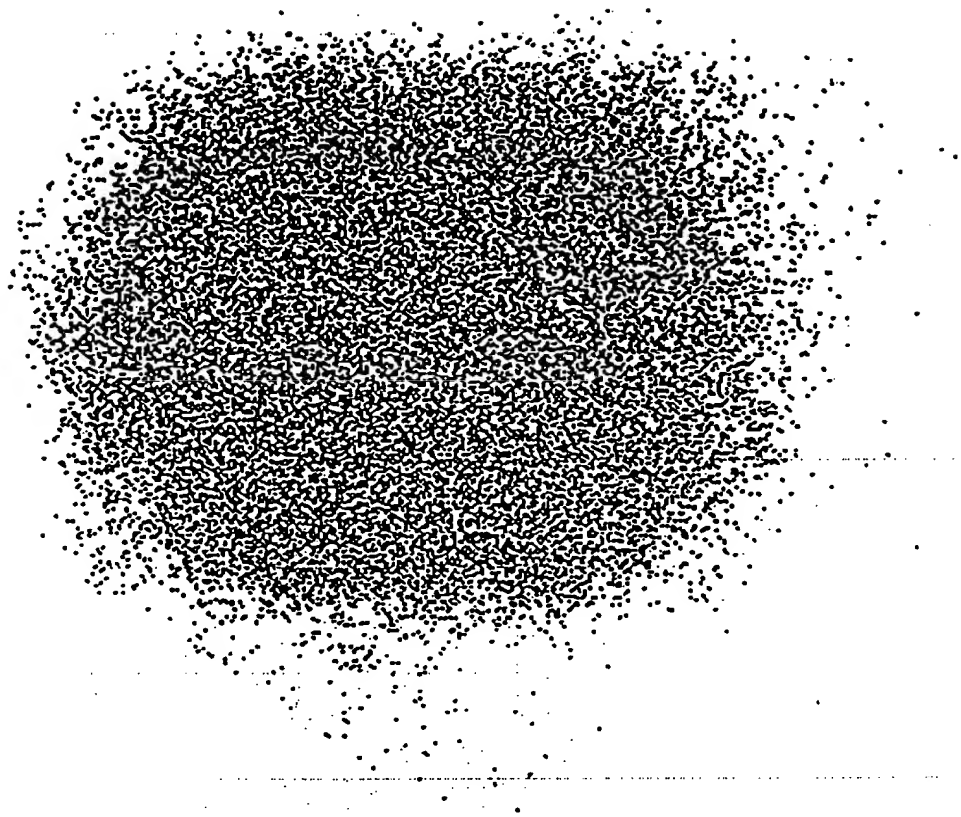


Without Hand-tuned Rotation Angles

—— inner cells of 1st pair of ASIN(k)

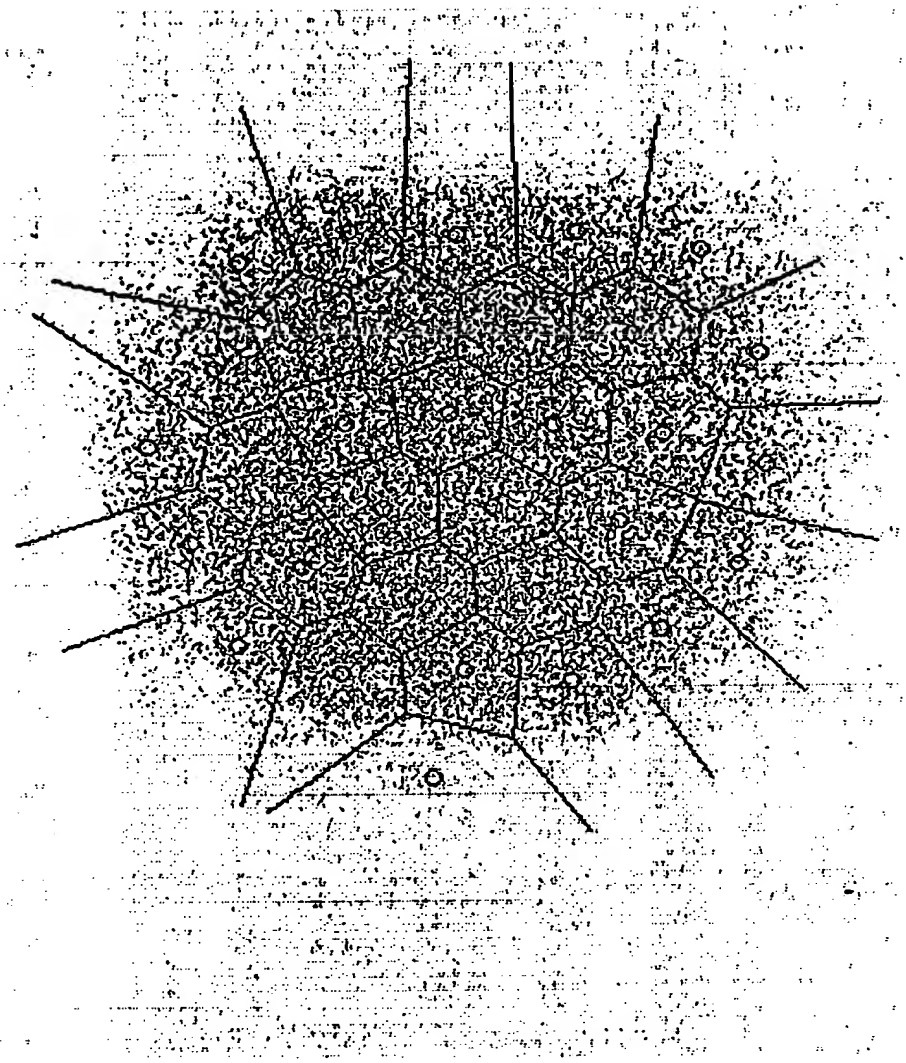
1st-stage VQ of

Fig. 25



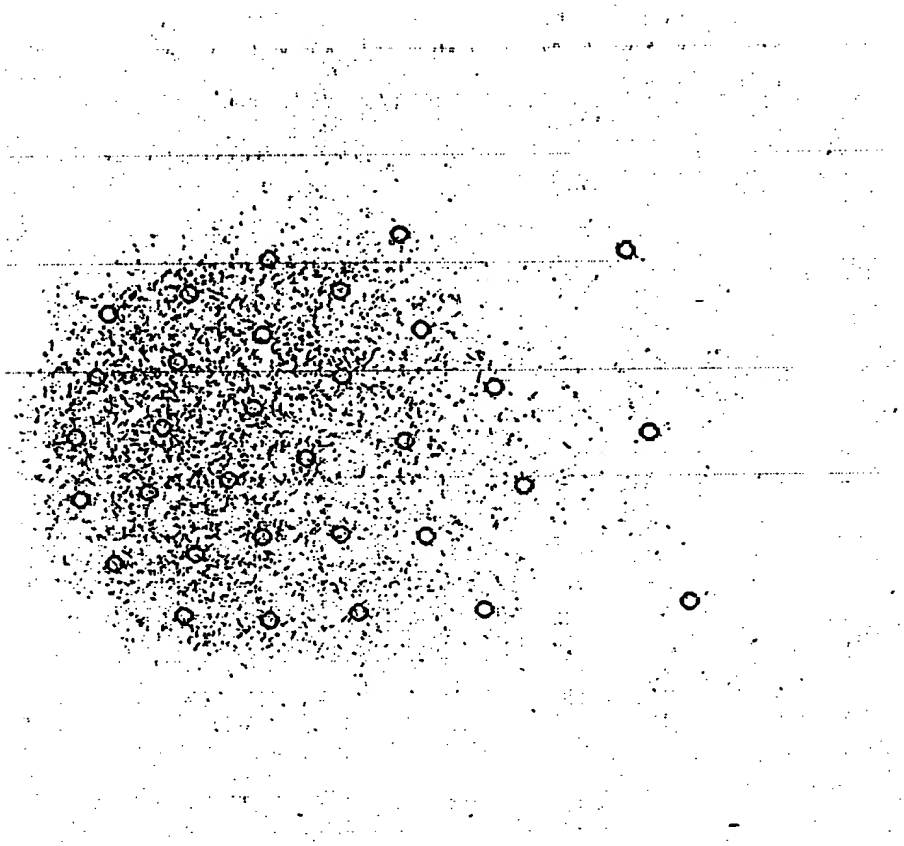
With hand-tuned rotation angles
— inner cells of 1st pare of ASIN(k)
of 1st-stage VQ of

Fig. 26



(hand tuning) Inner-cell 1st-stage VQ error vector distribution (gray dots) and corresponding 2nd-stage VQ codebook (small circles) for 1st pair of ASIN(k)

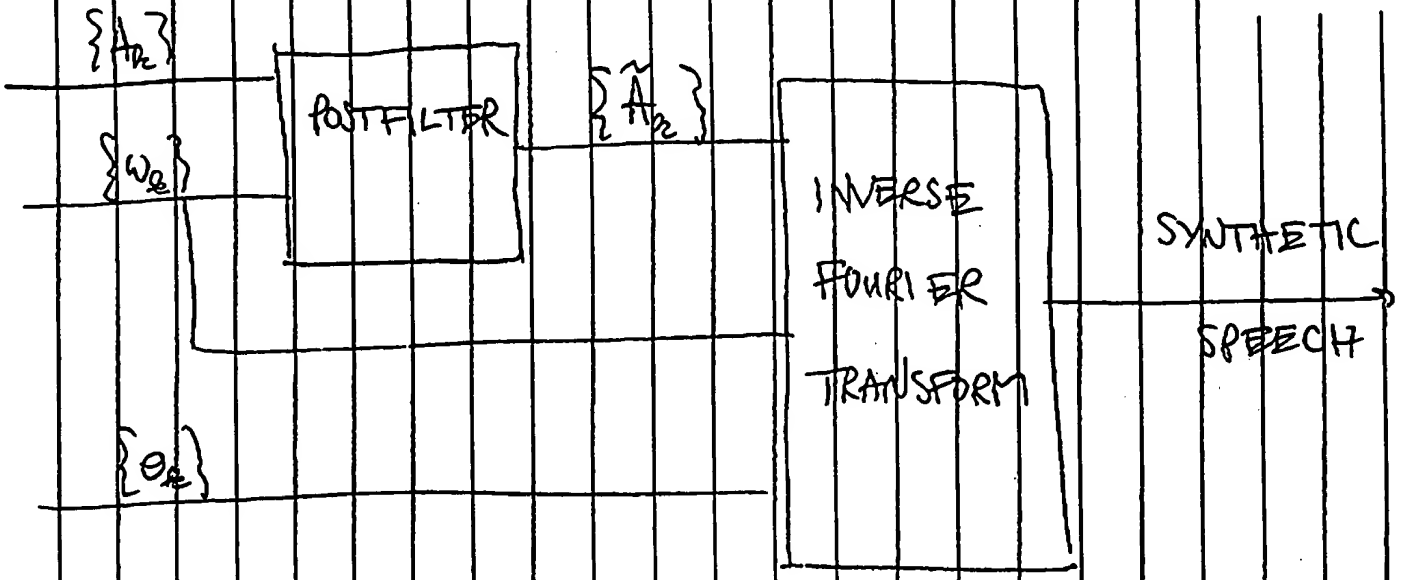
Fig. 27



Outer-cell 1st-stage VQ error vector distribution and corresponding 2nd-stage VQ codebook (small circles) for 1st pair of ASIN(k)

Fig. 28

SINUSOIDAL SPEECH SYNTHESIS



A_k = k^{th} sine-wave amplitude

ω_k = k^{th} sine-wave frequency

θ_k = k^{th} sine-wave phase

$\tilde{A}_k$ = k^{th} post-filtered sine-wave amplitude

Fig. 29

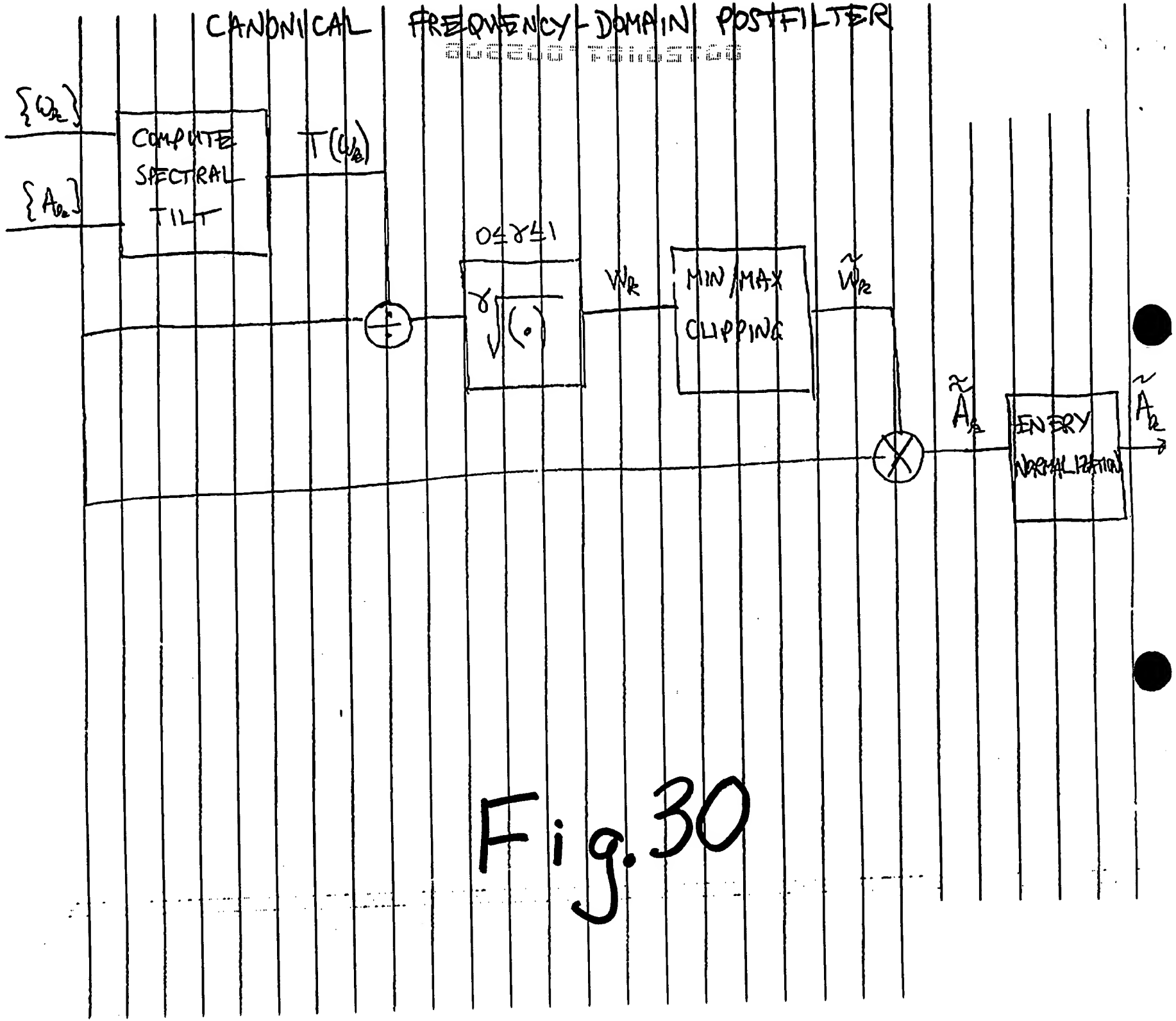

CANONICAL FREQUENCY-DOMAIN POSTFILTER
{ω_k}
{A_ω_k}
COMPUTE SPECTRAL TILT
T(ω_k)
0≤γ≤1
γ
W_k
MIN/MAX CLIPPING
W̃_k
Ã_k
ENERY NORMALIZATION
Ã_k

Fig. 30

CONSTANT COMPLEXITY POSTFILTER

$E(\omega)$ = amplitude envelope

ω_k = downsampled frequencies

$\tilde{E}(\omega)$ = post-filtered amplitude envelope

Fig. 31

CONSTANT COMPLEXITY POSTFILTER COMPUTED FROM CEPSTRAL COEFFICIENTS

C_m = m^{th} CEPSTRAL COEFFICIENT

$E(\omega)$ = AMPLITUDE ENVELOPE

$\tilde{\tilde{E}}(\omega)$ = POST-FILTERED AMPLITUDE ENVELOPE

Fig. 32

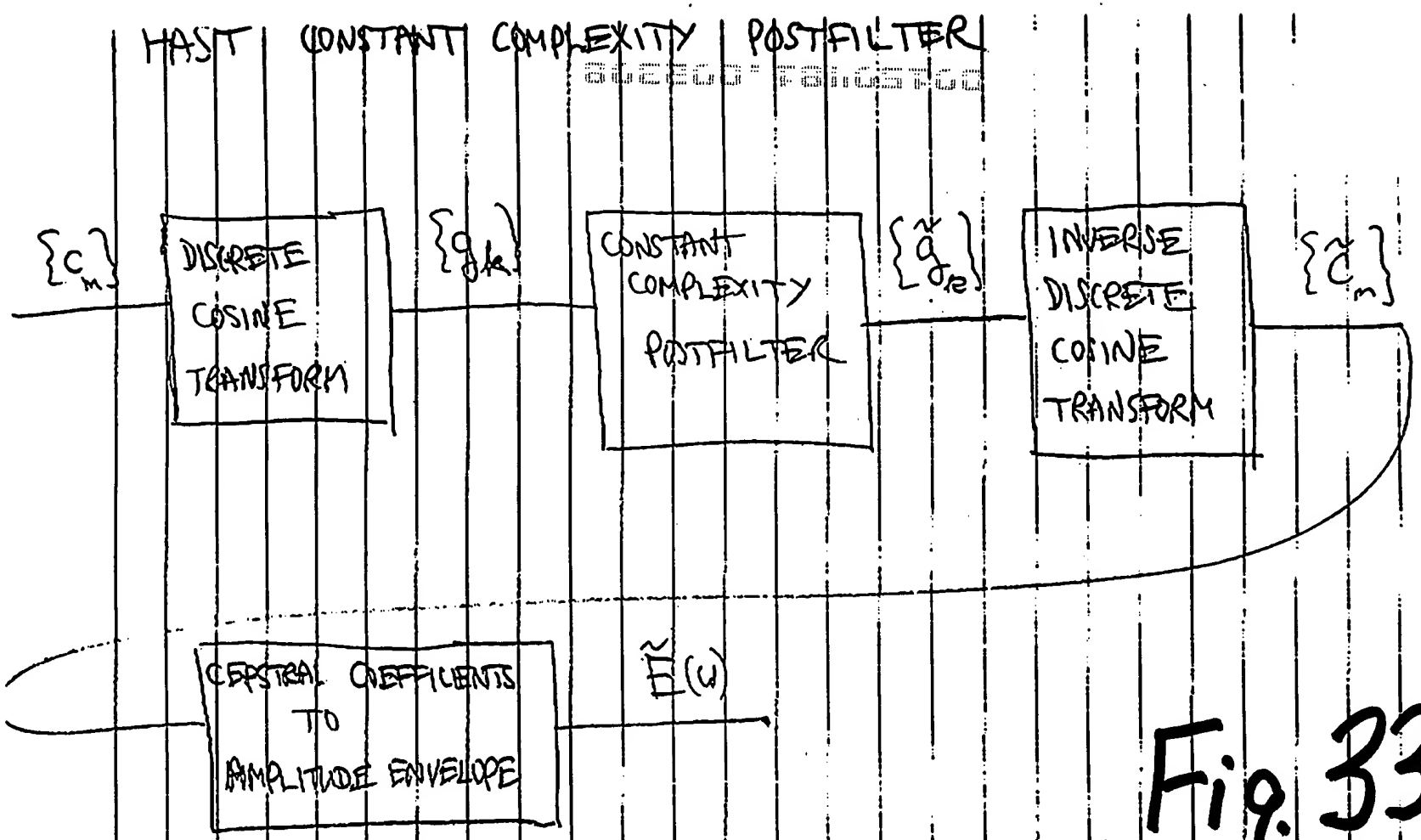


Fig. 33

c_m = m^{th} cepstral coefficient

g_{dk} = k^{th} DCT coefficient = k^{th} channel gain

$\tilde{g}_{dk}$ = k^{th} post-filtered channel gain

$\tilde{c}_m$ = m^{th} post-filtered cepstral coefficient

$\tilde{E}(\omega)$ = post-filtered amplitude envelope

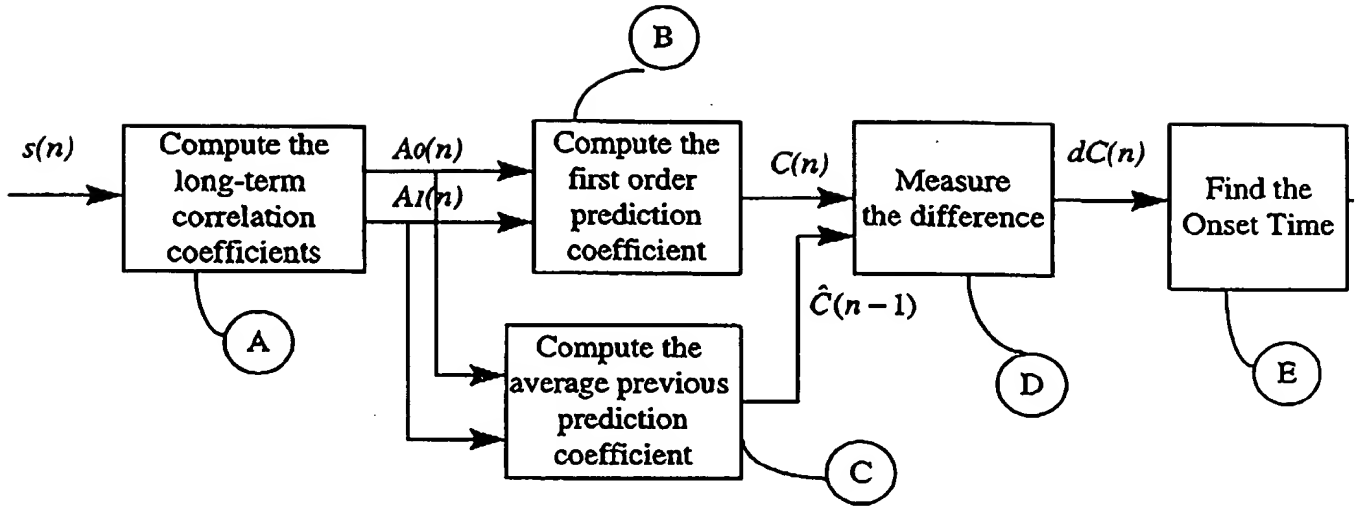


Figure 34.

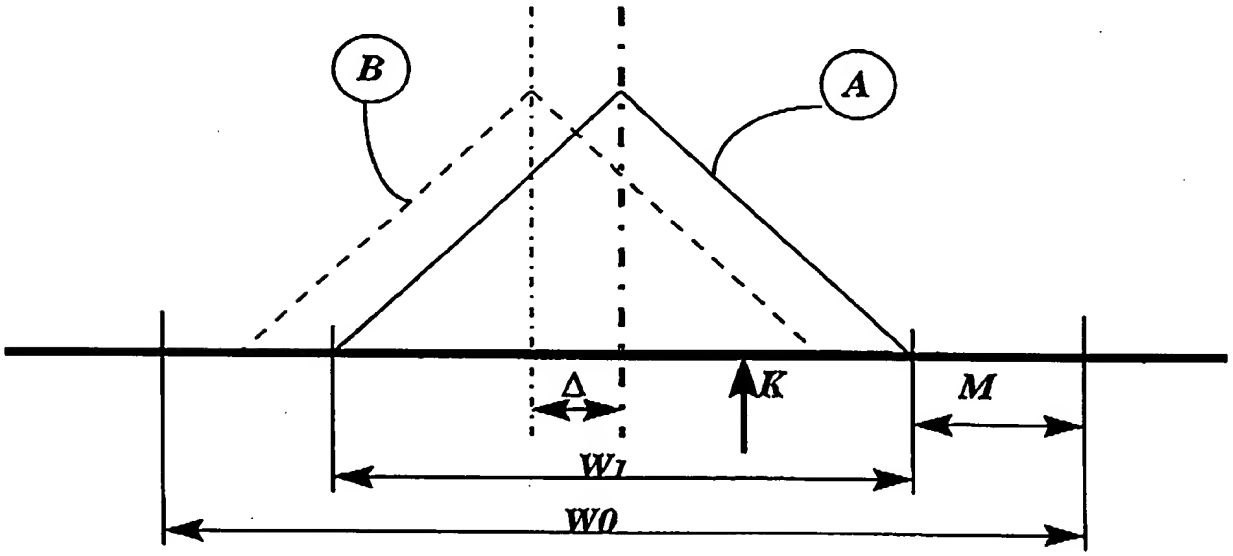


Figure 35.